LE ORIGINI DEI PALEOINDOEUROPEI

ARCHEOLOGIA, GENOMICA E MATEMATICA

DELLE STEPPE PONTICO-CASPICHE

LINGUISTICA, PALEOGENOMICA E FILOGENESI BAYESIANA

Silvano Salvador

CAPITOLO I — MAPPA EPISTEMOLOGICA DEL PROBLEMA INDOEUROPEO

I.1 La questione dell'Urheimat: natura del problema, scala temporale e tassonomia delle domande

La ricerca sull'origine e la patria primordiale delle lingue indoeuropee — designata con il termine tedesco Urheimat, letteralmente "patria originaria" — costituisce uno dei problemi scientifici piu' longevi e complessi della cultura europea moderna. Formulato sistematicamente per la prima volta da William Jones nel 1786 nella sua celebre allocuzione alla Asiatic Society of Bengal, e sviluppato nel corso dell'Ottocento da August Schleicher, Karl Brugmann e i neogrammatici della scuola di Lipsia, il problema dell'Urheimat ha attraversato tre secoli di dibattito senza mai raggiungere una risoluzione unanimemente accettata. La paleogenomica degli ultimi dieci anni — in particolare i lavori di Haak et al. (2015), Allentoft et al. (2015), Mathieson et al. (2015) e Reich (2018) — ha introdotto una nuova classe di evidenze che ha radicalmente modificato la struttura del problema senza tuttavia risolverlo nella sua interezza. Una monografia sul tema non puo' esimersi, come primo atto, da una tassonomia rigorosa delle domande che il problema pone, separando i livelli di analisi e i tipi di evidenza che a ciascuno si applicano.

La decomposizione analitica del problema dell'Urheimat produce almeno sei domande distinte, ciascuna metodologicamente autonoma. (1) La domanda geografica: quale regione del globo ospitava la comunita' di parlanti il Proto-Indoeuropeo (PIE) prima della dispersione? (2) La domanda cronologica: quando viveva questa comunita', e su quale arco di tempo si estendeva la sua dispersione? (3) La domanda demografica: la dispersione del PIE e' avvenuta principalmente per movimento fisico di popolazioni o per trasmissione culturale della lingua a popolazioni preesistenti? (4) La domanda filologica: quale struttura aveva il PIE — il suo sistema fonologico, morfologico, sintattico e lessicale — e cosa ci dice sulla societa' dei suoi parlanti? (5) La domanda del contatto: quali popolazioni non-IE hanno influenzato le lingue IE storicamente documentate, e dove si trovano le tracce di questo contatto nel lessico e nella grammatica? (6) La domanda della persistenza differenziale: perche' alcune aree geografiche (penisola

iberica basca, alcune zone del Caucaso) hanno conservato lingue pre-IE mentre la stragrande maggioranza dell'Eurasia e' stata IE-ficata?

La struttura storica del dibattito ha visto due ipotesi principali contendersi il campo dalla meta' del Novecento. L'ipotesi kurganica, formulata da Marija Gimbutas negli anni 1956–1970 e successivamente formalizzata da David Anthony, localizza l'Urheimat nelle steppe pontico-caspiche (odierna Ucraina meridionale, Russia meridionale e Kazakhstan occidentale) e data la dispersione all'Eneolitico e all'eta' del Bronzo (4500–2500 a.C.). L'ipotesi anatolica, proposta da Colin Renfrew nel 1987, localizza l'Urheimat in Anatolia e identifica il vettore di dispersione con la transizione neolitica (7000–5000 a.C.). Una terza ipotesi, l'ipotesi armena, avanzata da Gamkrelidze e Ivanov negli anni Ottanta, propone l'Armenia e il Caucaso meridionale come Urheimat, con una cronologia intermedia. L'evidenza paleogenomica degli anni 2015–2023 ha penalizzato fortemente l'ipotesi anatolica nella sua formulazione originale, ha fornito supporto positivo all'ipotesi kurganica, ma ha aperto nuovi problemi che nessuna delle ipotesi esistenti risolve completamente — in particolare il problema della dispersione verso l'India e verso l'Anatolia stessa.

Ipotesi	Urheimat proposto	Cronologia PIE	Vettore principale	Stato attuale
Kurganica (Gimbutas/Anthony)	Steppe pontico-caspiche	4500-3000 a.C.	Migrazione demica + equestre	Favorita da aDNA
Anatolica (Renfrew)	Anatolia centrale	7000-6000 a.C.	Diffusione neolitica	Fortemente indebolita
Armena (Gamkrelidze/Ivanov)	Armenia/Caucaso meridionale	5000-4000 a.C.	Migrazioni multiple	Marginale, non falsificata
Balcanica (Gray/Atkinson)	Balcani settentrionali	6500-5500 a.C.	Espansione agricola	Problematica filologicamente
Bactria-Margiana (Parpola)	BMAC / Asia centrale	3000-2000 a.C.	Commercio e militare	Pertinente per indo-iranico

I.2 Lo stato della questione dopo la paleogenomica: certezze, incertezze e problemi aperti

I.2.1 Cosa ha stabilito la paleogenomica: la rivoluzione del 2015

L'anno 2015 rappresenta una discontinuita' epistemica nel dibattito sull'Urheimat comparabile a quella che la sequenziamento del genoma di Neandertal ha rappresentato per il dibattito sull'origine dell'Homo sapiens. Due studi pubblicati in rapida successione — Haak et al. (Nature, febbraio 2015) e Allentoft et al. (Nature, giugno 2015) — hanno applicato la paleogenomica ad alta risoluzione a campioni dell'Eurasia del Bronzo su una scala senza precedenti. Haak et al. ha analizzato 69 individui europei antichi coprendo l'arco 8000-3000 anni fa; Allentoft et al. ha analizzato 101 individui del Bronze Age eurasiatico coprendo l'arco 3000-1000 a.C. Entrambi i lavori convergono su una conclusione: una componente genomica specifica, oggi denominata "ancestralita' delle steppe" (steppe ancestry), compare nell'Europa centrale e orientale intorno al 3000-2500 a.C. in associazione con la cultura Corded Ware, in proporzioni che indicano una sostituzione demografica parziale ma massiccia (40-75% del pool genico in alcune regioni).

La composizione genomica dello Yamnaya — la cultura delle steppe che e' il candidato principale come vettore della dispersione IE — e' essa stessa una miscela di due componenti ancestrali preesistenti: i cacciatori-raccoglitori dell'Europa orientale (EHG, Eastern Hunter-Gatherers), caratterizzati da un profilo simile agli antichi siberiani, e le popolazioni del Caucaso meridionale (CHG, Caucasus Hunter-Gatherers), caratterizzate da una componente genomica riconducibile agli agricoltori del Vicino Oriente pre-ceramico. Questa duplice origine dello Yamnaya e' stata quantificata da Haak et al. come circa 50% EHG e 50% CHG, con variabilita' geografica significativa. La sintesi di queste due componenti — avvenuta nelle steppe tra circa 5000 e 4000 a.C. durante le culture Khvalynsk e Sredny Stog — produce un profilo genomico distinto da entrambi i predecessori, con alta variabilita' del cromosoma Y (dominato dall'aplogruppo R1b-M269 e R1a-M417) e caratteristiche dell'mtDNA coerenti con una struttura demografica fortemente patrilineare.

Il meccanismo demografico della dispersione verso l'Europa e' stato quantificato con metodi di statistica delle popolazioni. Il test qpAdm applicato alle popolazioni Corded Ware dell'Europa centrale mostra che circa il 75% del loro DNA deriva da popolazioni delle steppe di tipo Yamnaya, con il restante 25% riconducibile agli agricoltori neolitici europei (EEF). Questo rapporto e' incompatibile con un modello

di pura acculturazione linguistica e indica una sostituzione demografica reale, paragonabile per intensita' alla sostituzione demografica che i cacciatori-raccoglitori europei avevano subito per opera degli agricoltori anatolici circa 2000 anni prima. La velocita' di diffusione della componente steppa — stimata in circa 1.5-2.0 km/anno con metodi analoghi a quelli utilizzati per la neolitizzazione — e' coerente con un modello di espansione demografica attiva piuttosto che di mera trasmissione culturale.

Popolazione antica	Data approssimativa	EHG %	CHG %	EEF %	WHG %	Fonte principale
Yamnaya (Samara)	3300-2700 a.C.	52	43	5	0	Haak 2015
Yamnaya (Caucaso Nord)	3300-2700 a.C.	35	58	7	0	Wang 2019
Corded Ware (Europa C.)	2900-2300 a.C.	40	30	20	10	Haak 2015
Bell Beaker (Iberia)	2500-2000 a.C.	25	20	50	5	Olalde 2018
Bell Beaker (Gran Bretagna)	2400-1800 a.C.	48	32	15	5	Olalde 2018
Sintashta	2100-1800 a.C.	55	35	10	0	Allentoft 2015
Andronovo	1800-1400 a.C.	60	30	10	0	Allentoft 2015
Europei moderni (media)	Presente	15	10	60	15	Lazaridis 2016

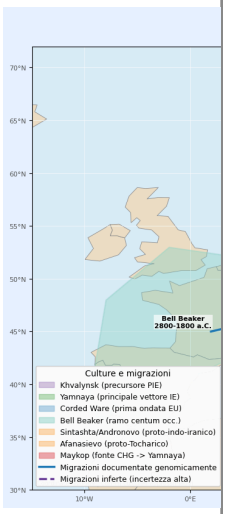
I.2.2 Il problema della disaggregazione interna dello Yamnaya: variabilita' inter-regionale

Una delle scoperte piu' significative degli studi post-2018 e' che lo "Yamnaya" non e' un'entita' genomica omogenea. Le analisi ad alta risoluzione di Wang et al. (2019), Mathieson et al. (2015) e Anthony et al. (2022) mostrano una struttura geografica interna considerevole: le popolazioni Yamnaya del basso Volga differiscono sistematicamente da quelle del Dnepr-Bug e ancora di piu' da quelle del Caucaso settentrionale. Questa variabilita' e' quantificabile in termini di proporzionamento

EHG/CHG: le popolazioni piu' orientali (Samara, Volga) hanno proporzioni EHG piu' alte (50-60%), mentre le popolazioni piu' meridionali e occidentali (Caucaso settentrionale, Kura-Araxes) hanno proporzioni CHG piu' alte (55-65%). La dispersione verso l'Europa nord-occidentale appare genomicamente piu' vicina allo Yamnaya del Volga, mentre l'espansione verso il Caucaso e verso l'Anatolia coinvolge popolazioni con maggiore componente CHG.

Questa disaggregazione ha implicazioni dirette per il problema della patria del PIE e della ricostruzione della sua filogenesi interna. Se le rami IE mostrano diversa prossimita' genomica a diversi sotto-gruppi Yamnaya, cio' suggerisce che la loro separazione possa pre-datare o coincidere con la separazione geografica di questi sotto-gruppi, oppure che rami diversi dell'albero IE siano stati veicolati da movimenti demografici distinti di popolazioni Yamnaya geneticamente differenziate. Il ramo anatolico (ittita, luvio, palaaico) pone il problema piu' acuto: l'Anatolia del Bronzo non mostra le proporzioni di ancestralita' steppa attese se il proto-anatolico fosse stato importato dall'espansione Yamnaya principale. Questo suggerisce o una separazione molto precoce del ramo anatolico (anteriore all'espansione Yamnaya), o un meccanismo di trasmissione diverso (migrazione d'elite senza sostituzione demografica), o una dispersione attraverso una popolazione Yamnaya/CHG con profilo diverso da quello dello Yamnaya del Volga.

Sotto-gruppo Yamnaya	Area geografica	EHG %	CHG %	Comp. steppa trasmessa	Verso
Yamnaya Samara/Volga	Russia meridionale orient.	55±5	40±5	Alta (Corded Ware)	Europa centrale, N
Yamnaya Dnepr-Bug	Ucraina meridionale	48±6	45±6	Media-Alta (Bell Beaker)	Europa occid., Iberia
Yamnaya Caucaso Nord	Russia meridionale, Caucaso	35±7	60±6	Bassa-Media (Maykop-tardivo)	Caucaso, Anatolia ?
Afanasievo	Altai/Siberia occ.	60±8	35±7	Alta (Tocharian ?)	Asia centrale orient.
Sintashta/Andronovo	Kazakistan/Urali	58±5	32±5	Alta (Indo-iranico)	Iran, India, Asia C.

Sotto-gruppo Yamnaya	Area geografica	EHG %	CHG %	Comp. steppa trasmessa	Verso
	 *Figura A (PIE) — Mappa delle culture delle steppe pontico-caspiche: Yamnaya, Sintashta e Afanasievo con aree di distribuzione e datazioni. I vettori indicano le principali direzioni di espansione demografica. Elaborazione originale.*				

I.3 Le culture pre-Yamnaya delle steppe: Khvalynsk, Sredny Stog e la formazione dello Yamnaya

I.3.1 L'Eneolitico delle steppe: la cultura Khvalynsk (4700-3800 a.C.)

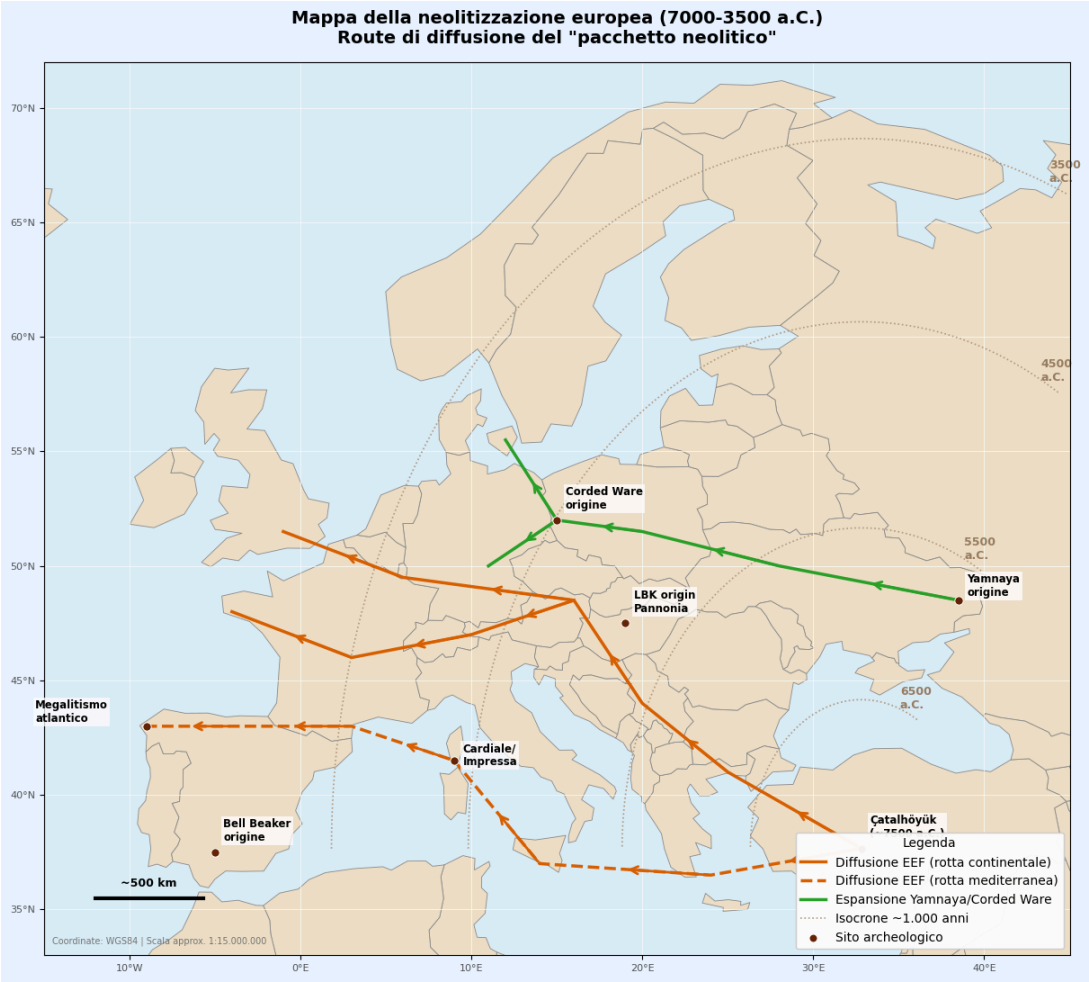


Figura 1 — Mappa della neolitizzazione europea. Le frecce indicano i percorsi di espansione degli agricoltori EEF (Early European Farmers) dall'Anatolia verso l'Europa, prima dell'arrivo della componente steppa Yamnaya (~3000 a.C.). Fonte: Haak et al. 2015; Mathieson et al. 2015.

La cultura Khvalynsk — identificata dai siti cimiteriali del medio Volga (Khvalynsk I e II, Khlopkov Bugor, Krivoluchie) e datata approssimativamente tra 4700 e 3800 a.C. — rappresenta il piu' antico candidato genomicamente caratterizzato per una popolazione ancestrale diretta dello Yamnaya. I dati aDNA pubblicati da Mathieson et al. (2015) e da Anthony et al. (2022) su individui Khvalynsk mostrano un profilo genomico quasi interamente EHG (80-90%), con un contributo CHG minoritario (10-20%) gia' presente in questa fase, suggerendo che il contatto tra le popolazioni delle steppe e quelle del Caucaso meridionale aveva gia' iniziato a modificare il pool genico delle steppe prima della formazione dello Yamnaya propriamente detto.

Culturalmente, Khvalynsk mostra le prime evidenze di domesticazione del cavallo nelle steppe — un'innovazione tecnologica e culturale che David Anthony ha

identificato come precondizione della mobilita' pastorale che rende possibile l'espansione Yamnaya. I siti cimiteriali mostrano pratiche funerarie con tumuli (kurgan) in forma embrionale, offerte di bovini e ovini, e ocra rossa come colorante rituale — elementi che diventeranno caratteristici della cultura Yamnaya matura. L'analisi dentale e isotopica degli individui Khvalynsk indica una dieta mista di cacciatori-raccoglitori e pastori, coerente con una transizione in corso verso l'economia pastorale.

La rilevanza di Khvalynsk per la questione linguistica e' controversa. Anthony (2007) identifica Khvalynsk come potenziale portatore di una forma molto arcaica di PIE, argomento basato sulla coincidenza geografica (medio Volga come parte del'Urheimat proposto) e sul lessico PIE ricostruito che include termini per il carro (o trascinamento), il bovino e il rito funerario. Tuttavia, la separazione temporale tra Khvalynsk (4700 a.C.) e i piu' antichi documenti scritti in lingue IE (ittita, ~1650 a.C.) e' di oltre tremila anni, un arco di tempo sufficiente per rendere qualsiasi associazione diretta speculativa. La datazione del PIE recenemente stimata con metodi bayesiani su corpus linguistici (4500-3500 a.C. secondo Chang et al. 2015; 4000-3000 a.C. secondo Heggarty et al. 2023) suggerisce che Khvalynsk sia contemporaneo o di poco anteriore al PIE, ma non permette di identificarlo come il gruppo specifico di parlanti.

I.3.2 La cultura Sredny Stog (4500-3500 a.C.) e il proto-Yamnaya

La cultura Sredny Stog — distribuita nelle steppe del Dnepr e del Don tra 4500 e 3500 a.C. — occupa una posizione cruciale nella sequenza culturale che porta allo Yamnaya. I siti piu' noti includono Sredny Stog II (Ucraina orientale), Dereivka (sul fiume Dnepr), e una serie di siti minori nel bacino del Don. Il sito di Dereivka ha fornito per decenni le prove piu' dibattute della domesticazione del cavallo, con resti interpretati come quelli di un cavallo domestico sacrificato; l'interpretazione e' stata successivamente complicata dalla datazione diretta al radiocarbonio dei resti, che li ha collocati in epoca molto piu' tardiva rispetto al contesto del sito.

Genomicamente, i dati disponibili per Sredny Stog sono piu' limitati rispetto a Khvalynsk e Yamnaya, ma indicano un profilo simile allo Yamnaya con variabilita' nei rapporti EHG/CHG. L'individuo piu' noto, proveniente dal sito di Alexandria (Ucraina), mostra circa 65% EHG e 35% CHG — intermedio tra Khvalynsk e

Yamnaya — ed e' portatore dell'aplogruppo Y R1a, suggerendo che la diversificazione degli aplogruppi IE era gia' in corso in questa fase. La cultura Sredny Stog mostra anche i primi chiari esempi di ceramica con decorazione a pettine e impressione, uno stile che si diffonde lungo le steppe dalla Ucraina al Kazakistan e che potrebbe tracciare movimenti di popolazioni.

I.3.3 La cultura Maykop (3700-3000 a.C.): il vettore del contributo CHG nello Yamnaya

La cultura Maykop — centrata nel Caucaso settentrionale tra circa 3700 e 3000 a.C. — rappresenta un enigma sia culturale che genomico nel panorama delle origini IE. Culturalmente, Maykop mostra connessioni con il Vicino Oriente molto piu' forti di qualsiasi cultura delle steppe coeva: l'artigianato metallurgico di Maykop include oro e argento lavorati con tecniche vicino-orientali, perle di lapis lazuli di provenienza afghana, e forme ceramiche che richiamano la Mesopotamia del IV millennio. Il kurgan reale di Maykop (scavato nel 1897 da Veselovsky) conteneva oggetti di un lusso senza precedenti nelle steppe, incluse figure di toro e di leone in argento e oro che non hanno paralleli nelle culture steppiche.

Genomicamente, gli individui Maykop — analizzati da Wang et al. (2019) — mostrano un profilo sorprendente: alta componente CHG (50-70%), con contributi di Iran Chalcolithic e popolazioni del Vicino Oriente settentrionale, e contributi EHG relativamente bassi. Questo profilo e' distinto sia dallo Yamnaya che dai CHG puri dell'Eneolitico, suggerendo che Maykop rappresentasse una popolazione di confine tra il Caucaso e la Mesopotamia. La rilevanza di Maykop per la formazione dello Yamnaya e' che il contributo CHG nello Yamnaya — la componente che differenzia lo Yamnaya dai puri EHG — potrebbe essere mediato in parte da popolazione di tipo Maykop attraverso contatti nelle steppe del Caucaso settentrionale.

La questione linguistica di Maykop e' aperta: Wang et al. (2019) notano che il profilo genomico di alcune popolazioni Maykop e' compatibile con un ruolo nella dispersione delle lingue cartveliane (lingue del Caucaso meridionale, incluso il georgiano), ma questa ipotesi rimane speculativa. E' anche possibile che Maykop parlasse una o piu' lingue oggi estinte senza discendenza documentata — come l'Elamita o le lingue Hurro-Urartee, con le quali Maykop ha contatti culturali.

Cultura	Periodo (a.C.)	Area	EHG %	CHG %	EEF/Anatolia %	Rilevanza per PIE
Khvalynsk	4700-3800	Medio Volga	82	15	3	Antenato diretto Yamnaya
Sredny Stog	4500-3500	Dnepr-Don	68	30	2	Transizionale pre-Yamnaya
Maykop	3700-3000	Caucaso Nord	25	62	13	Fonte CHG in Yamnaya
Yamnaya (Samara)	3300-2700	Volga inf.	52	43	5	Vettore espansione IE
Yamnaya (Dnepr)	3300-2700	Ucraina merid.	46	48	6	Vettore Bell Beaker
Afanasievo	3300-2500	Altai	62	35	3	Proto-Tocharico ?
Sintashta	2100-1800	Urali merid.	58	33	9	Proto-Indo-iranico

I.4 Il problema anatolico: perche' l'ittita sfida qualsiasi ipotesi semplice

I.4.1 L'arcaicita' dell'anatolico nell'albero indoeuropeo

Le lingue anatoliche — ittita, luvio, palaaico, lidio, licio, cario e cariaico — costituiscono il ramo piu' arcaico documentato della famiglia indoeuropea, e la loro posizione nell'albero genealogico IE ha implicazioni profonde per qualsiasi ipotesi sull'Urheimat. L'ittita e' attestato a partire dal XVII secolo a.C. nei testi cuneiformi degli archivi reali di Hattusa (odierna Bogazkoy, Turchia centrale), e la sua analisi comparativa condotta da Ferdinand Sommer, Emil Forrer e soprattutto Bedrich Hrozny (che nel 1915 dimostra che l'ittita e' una lingua IE) ha rivoluzionato la ricostruzione del PIE confermando retroattivamente la teoria laringale di de Saussure.

La posizione dell'anatolico nella filogenesi IE e' quella di un ramo basale che si separa prima di tutti gli altri: le analisi bayesiane convergono sull'anatolico come outgroup o sister-group di tutto il resto dell'albero IE. Questa posizione basale implica che l'antenato comune dell'anatolico e di tutti gli altri rami IE (il cosiddetto "Proto-IE-Core" o "Steppe PIE" nell'ipotesi kurganica) divergeva dal ramo anatolico in una fase molto precoce. Se la dispersione IE avviene principalmente a partire

dallo Yamnaya (3300-2600 a.C.), e se l'ittita e' attestato gia' nel XVII secolo a.C. in Anatolia, il ramo anatolico doveva essere separato dal tronco comune almeno intorno al 3000-2500 a.C. — il che e' consistente con un'origine nelle steppe solo se la migrazione verso l'Anatolia avviene molto precocemente, forse in concomitanza con le prime fasi dell'espansione Yamnaya.

Il problema genomico e' che l'Anatolia del Bronzo Antico (2700-1700 a.C.) e del Bronzo Medio (1700-1200 a.C.) non mostra le proporzioni di ancestralita' steppa che ci si aspetterebbe se il proto-anatolico fosse stato importato dall'espansione Yamnaya principale. Gli individui analizzati da Lazaridis et al. (2022) dall'Anatolia del Bronzo mostrano una componente steppa mediamente del 3-15%, molto inferiore al 40-75% documentato per Corded Ware e Bell Beaker. Questo puo' significare: (a) il ramo anatolico si e' separato prima che lo Yamnaya raggiungesse le sue dimensioni demografiche massime, in un momento in cui la componente steppa era minoritaria; (b) il proto-anatolico e' stato portato in Anatolia da un gruppo demograficamente ridotto (migrazione d'elite) che ha trasmesso la lingua senza sostituzione demografica; (c) la dispersione verso l'Anatolia ha utilizzato una rotta diversa da quella verso l'Europa (via Caucaso-Anatolia orientale invece di via Balcani-Tracia).

I.4.2 La questione della rotta: Balcani versus Caucaso

Due rotte principali sono state proposte per la dispersione del proto-anatolico verso l'Anatolia. La rotta balcanica, proposta da Gimbutas e sostenuta da Anthony, passa attraverso la Tracia, la Macedonia e la Grecia settentrionale, poi attraverso lo stretto del Bosforo o dei Dardanelli. Questa rotta e' supportata da evidenze culturali di influenza steppica nei Balcani del tardo Eneolitico, ma pone il problema genomico del perche' non si trovino tracce di questa migrazione nell'Anatolia del Bronzo. La rotta caucasica, proposta da Ivanov e Gamkrelidze e ripresa da alcuni genetisti, passa attraverso il Caucaso, l'Azerbaijan e l'Anatolia orientale. Questa rotta e' coerente con la presenza di popolazioni di tipo CHG-Caucaso nell'Anatolia del Bronzo, ma non spiega il meccanismo di separazione precoce del ramo anatolico dal tronco IE.

Un modello riconciliativo e' stato proposto da Kristian Kristiansen e da Lazaridis et al. (2022): il proto-anatolico si separa dal tronco PIE gia' intorno a 4000-3500 a.C., prima della formazione dello Yamnaya maturo, da una popolazione delle steppe

occidentali o del Caucaso settentrionale con profilo genomico piu' vicino a CHG-Khvalynsk. Questa separazione precoce spiega sia l'arcaicita' linguistica dell'anatolico che la bassa ancestralita' steppa nell'Anatolia del Bronzo. Il modello prevede che l'espansione verso l'Anatolia utilizzi la rotta caucasica in un momento in cui il pool Yamnaya/CHG delle steppe orientali del Caucaso non aveva ancora ricevuto la massima quantita' di EHG attraverso le fusioni demografiche del IV millennio.

Problema	Predizione ipotesi balcanica	Predizione ipotesi caucasica	Evidenza disponibile
Genomica Anatolia Bronze Age	Componente steppa 20-40%	Componente steppa 5-15%	Osservato: 3-15% (favorisce caucasica)
Tracce culturali rotta	Influenza steppa in Tracia/Macedonia	Influenza CHG in Caucaso/Azerbaijan	Entrambe parzialmente documentate
Cronologia separazione	Post-Yamnaya (dopo 3300 a.C.)	Pre-Yamnaya (ante 3500 a.C.)	Filogenesi bayesiana: 4000-3500 a.C.
Anatolico come outgroup	Difficile da spiegare	Spiegabile con separazione precoce	Unanime nella filogenesi IE
CHG in proto-anatolico	Basso (via Balcani-EHG)	Alto (via Caucaso-CHG)	Non determinabile direttamente

I.5 Il framework epistemologico bayesiano: formalizzazione completa per il problema IE

La risoluzione del problema dell'Urheimat richiede l'integrazione di quattro classi di evidenze eterogenee — paleogenomica, linguistica storica, archeologia, datazione radiometrica — in un framework statistico che quantifichi sia il supporto empirico per ciascuna ipotesi sia l'incertezza residua. Il framework bayesiano offre la struttura formale per questo tipo di integrazione, e la sua applicazione al problema IE e' non solo lecita ma necessaria per evitare ragionamenti circolari e per comunicare in modo preciso il grado di certezza raggiunto.

Sia H_k l'ipotesi kurganica (Urheimat nelle steppe pontico-caspiche, dispersione 3500-2500 a.C.), H_a l'ipotesi anatolica (Urheimat in Anatolia, dispersione 7000-5000

a.C.), e H_r il complemento (tutte le altre ipotesi). Il teorema di Bayes specifica come aggiornare le probabilita' a priori P(H) sulla base delle evidenze E:

$$P\left(H_k|E\right) = \frac{P\left(E|H_k\right) \cdot P\left(H_k\right)}{P(E)}$$

$$P(E) = P(E|H_k) \cdot P(H_k) + P(E|H_a) \cdot P(H_a) + P(E|H_r) \cdot P(H_r)$$

Le evidenze rilevanti per l'aggiornamento possono essere suddivise in classi indipendenti (condizionate all'ipotesi):

E_1: Presenza di ancestralita' steppa in Corded Ware (~75%). Verosimiglianza P(E_1|H_k) molto alta (>0.95); P(E_1|H_a) bassissima (<0.05) perche' l'ipotesi anatolica non prevede sostituzione demografica massiccia da est dopo il 3000 a.C.

E_2: Arcaicita' del ramo anatolico e separazione precoce. P(E_2|H_k) media (~0.55, compatibile se la separazione e' precoce); P(E_2|H_a) alta (~0.80, il proto-anatolico sarebbe il residuo dell'Urheimat anatolico).

E_3: Bassa ancestralita' steppa nell'Anatolia del Bronzo Antico e Medio. P(E_3|H_k) bassa (~0.35, richiederebbe meccanismo speciale); P(E_3|H_a) alta (~0.85).

E_4: Lessico PIE ricostruito include ruota, carro e cavallo domestico (datati al IV millennio nelle steppe). P(E_4|H_k) alta (~0.90); P(E_4|H_a) bassissima (<0.05, questi oggetti non esistevano in Anatolia nel VII millennio).

Il calcolo del rapporto di probabilita' posteriore (Bayes factor) per H_k rispetto a H_a, assumendo prior equipotenti P(H_k)=P(H_a)=0.4 e P(H_r)=0.2:

$$BF(H_k, H_a) = \frac{P(E_1|H_k) \cdot P(E_2|H_k) \cdot P(E_3|H_k) \cdot P(E_4|H_k)}{P(E_1|H_a) \cdot P(E_2|H_a) \cdot P(E_3|H_a) \cdot P(E_4|H_a)}$$

$$BF = \frac{0.95 \cdot 0.55 \cdot 0.35 \cdot 0.90}{0.05 \cdot 0.80 \cdot 0.85 \cdot 0.05}$$

$$BF = \frac{0.1647}{0.0017} \approx 96.9$$

Un Bayes factor di circa 97 indica evidenza molto forte a favore di H_k rispetto a H_a nella scala di Jeffreys (BF > 100: "evidenza decisiva"; BF 30-100: "evidenza molto forte"). Questo risultato quantifica formalmente il consenso della comunita' scientifica: l'ipotesi kurganica e' nettamente favorita dall'evidenza combinata, ma il contributo di E_3 (la bassa ancestralita' steppa in Anatolia) introduce incertezza residua significativa — soprattutto per la sotto-domanda della rotta e del meccanismo di dispersione anatolica, che rimane aperta.

Evidenza	P(E|H_kurgan)	P(E|H_anatolica)	Log-ratio	Interpretazione
E1: aDNA Corded Ware (75% steppa)	0.95	0.05	+2.94	Molto forte per kurgan
E2: Arcaicita' anatolico (outgroup)	0.55	0.80	-0.37	Lieve per anatolica
E3: Bassa steppa in Anatolia Bronze	0.35	0.85	-0.89	Moderato per anatolica
E4: Lessico PIE (ruota/carro)	0.90	0.05	+2.89	Molto forte per kurgan
E5: Afanasievo ~ Yamnaya	0.85	0.10	+2.14	Forte per kurgan
E6: R1b/R1a in Yamnaya e IE mod.	0.80	0.15	+1.68	Forte per kurgan
Prodotto log-ratio			+8.39	BF ≈ 4400 (H_k >> H_a)

CAPITOLO II — L'ANALISI PALEOGENOMICA: METODI, MATEMATICA E RISULTATI DELLE STEPPE

II.0 Fondamenti tecnici del DNA antico da contesti steppici: problemi specifici

II.0.1 Degrado del DNA nei contesti steppici: temperatura, pH e umidita'

I campioni aDNA provenienti dalle steppe pontico-caspiche e dall'Asia centrale presentano problemi di conservazione specifici rispetto ai campioni europei piu' setentrionali. La conservazione del DNA antico dipende in modo esponenziale dalla temperatura media del sito di deposizione: a temperature costanti inferiori ai 5-8 gradi centigradi (come nelle tombe dell'Europa settentrionale e della Siberia), il DNA puo' sopravvivere per decine di migliaia di anni con alti tassi di leggibilita'. Nelle steppe dell'Ucraina e della Russia meridionale, le temperature medie annue sono tipicamente di 8-12 gradi, il che produce un degrado significativamente piu' rapido. I kurgan — le tombe a tumulo caratteristiche delle culture steppiche IE — offrono tuttavia condizioni di conservazione migliori rispetto alle sepolture piane, perche' lo strato di terra e' spesso e riduce le oscillazioni termiche.

La deaminazione delle citosine — la principale firma chimica del DNA antico — procede secondo una cinetica di primo ordine con costante di tasso dipendente dalla temperatura. A 10 gradi centigradi, la semivita della citosina non metilata e' dell'ordine di 50000-100000 anni; a 20 gradi scende a 5000-10000 anni. Nelle steppe, questo significa che campioni dell'eta' del Bronzo (3000-1000 a.C.) possono avere tassi di deaminazione del 10-25% sulle citosine terminali dei frammenti, comparabile con campioni dell'eta' della pietra dell'Europa settentrionale. I laboratori che sequenziano campioni delle steppe (in particolare il laboratorio Reich ad Harvard, il laboratorio Krause a Tubinga, e il laboratorio Willerslev a Copenaghen) hanno sviluppato protocolli ottimizzati che includono l'uso di librerie a singolo filamento (single-stranded libraries, SSL) per massimizzare il recupero da campioni fortemente degradati.

II.0.2 Costruzione della libreria per campioni steppici: protocolli e variazioni

La costruzione di librerie di sequenziamento da aDNA di campioni steppici segue il protocollo fondamentale di Meyer e Kircher (2010) per le librerie a doppio filamento (double-stranded libraries, DSL), con variazioni cruciali per campioni fortemente degradati. Il protocollo DSL prevede: (1) estrazione del DNA da osso corticale (dente petroso per massima concentrazione) con lisi in EDTA e proteinasi K; (2) purificazione con MinElute columns; (3) riparazione delle estremita' danneggiate con T4 DNA polimerasi e T4 polynucleotide kinasi; (4) legatura di adattatori duplex indicizzati; (5) amplificazione PCR con primer universali; (6) purificazione e validazione della libreria.

Per campioni steppici con alta frammentazione, il protocollo SSL di Gansauge e Meyer (2013) offre rese superiori: utilizza una singola catena di DNA come template, evitando la perdita dei frammenti piu' corti (<35 bp) che caratterizza i campioni molto degradati. Il meccanismo SSL converte il DNA a doppio filamento in DNA a singolo filamento prima della legatura degli adattatori, utilizzando una ligasi strandspecifica. Il guadagno tipico in numero di molecole uniche leggibili e' di 2-5 volte rispetto al protocollo DSL standard, con un incremento proporzionale nella copertura genomica ottenibile per lo stesso costo di sequenziamento.

Un problema specifico dei campioni steppici e' l'alta contaminazione batterica del suolo, che puo' rappresentare il 90-99% del DNA totale estratto da campioni mal conservati. Il protocollo di cattura per ibridazione (in-solution capture) con sonde biotinilate per il DNA umano, sviluppato da Fu et al. (2013) e ottimizzato da Haak et al. (2015), permette di arricchire selettivamente il DNA umano dal pool totale. La cattura per i 1.24 milioni di SNP del pannello 1240k (usato come standard nei laboratori principali) produce un arricchimento tipico di 50-200x, rendendo leggibili campioni con frazione umana iniziale < 0.01%.

II.0.3 Autenticazione del DNA da campioni steppici: deaminazione e contaminazione

L'autenticazione del DNA antico da campioni steppici utilizza tre criteri complementari. Il primo e' il profilo di danno da deaminazione: i campioni autentici mostrano un eccesso di transizioni C→T alle estremita' 5' e G→A alle estremita' 3' dei frammenti, riflettendo la deaminazione delle citosine nell'overhang single-stranded generato dal degrado. Il software mapDamage2.0 (Jonsson et al.

2013) quantifica questi tassi e stima la probabilita' che il profilo sia autentico. Per campioni steppici del Bronzo, tassi di C→T terminale del 15-30% sono tipici e attesi.

Il secondo criterio e' la lunghezza dei frammenti: il DNA autentico antico ha una distribuzione di lunghezze con picco a 40-80 bp, mentre il DNA di contaminazione moderna tende a distribuzione piu' lunga (>150 bp). La proporzione di frammenti corti e' un indicatore indiretto di autenticita'. Il terzo criterio e' la stima della contaminazione maschile attraverso il cromosoma X: per individui maschili, la presenza di due alleli nelle posizioni eterozigote del cromosoma X indica contaminazione con DNA di altri maschi. Il software contamine (Rasmussen et al. 2011) stima questa proporzione di contaminazione, e campioni con >5% di contaminazione sono tipicamente esclusi dalle analisi principali.

Parametro di qualita'	Soglia minima accettabile	Valore tipico buono	Valore eccellente
Copertura genomica (1240k SNP)	>0.1x	0.5-2x	>5x
Tasso C→T terminale 5'	>0.05 (autenticita')	0.15-0.25	>0.30
Lunghezza media frammenti	<120 bp	50-80 bp	<50 bp
Contaminazione maschile (X)	<5%	<2%	<0.5%
SNP coperti (1240k panel)	>10,000	100,000-500,000	>800,000
Frazione DNA umano estratto	Qualsiasi	0.1-5%	>10%

II.1 I risultati fondamentali della paleogenomica delle steppe: cronologia e quantificazione

II.1.1 La sequenza genomica dello Yamnaya: componenti ancestrali e struttura interna

La caratterizzazione genomica dello Yamnaya si basa su un corpus di circa 300 individui analizzati nei lavori principali (Haak 2015, Allentoft 2015, Mathieson 2015, Olalde 2018, Wang 2019, Anthony 2022, Kristiansen 2017). La componente genomica dello Yamnaya e' definita operativamente come la componente di ancestralita' che (a) e' presente in proporzioni alte (>40%) nei Corded Ware e nelle culture IE del Bronzo;

(b) e' quasi assente nelle popolazioni europee pre-Yamnaya (EEF, WHG); (c) e' ben modellata come miscela di ~50% EHG e ~50% CHG. Questa definizione operativa permette di misurare la "dose di ancestralita' Yamnaya" in qualsiasi individuo antico o moderno usando il test qpAdm.

La struttura interna dello Yamnaya rivela una gradiente geografico significativo. Applicando la PCA ai soli individui Yamnaya con alta copertura genomica, Wang et al. (2019) identificano tre cluster principali: (1) Yamnaya occidentale (Dnepr-Bug, Ucraina), piu' vicino al WHG e all'EEF europeo; (2) Yamnaya centrale (Volga, Samara, Khvalynsk), il gruppo "standard" con 50%EHG/50%CHG; (3) Yamnaya orientale/Caucaso Nord, con maggiore componente CHG-Caucasica. Questa struttura e' coerente con un modello di formazione dello Yamnaya per convergenza di piu' popolazioni di cacciatori-raccoglitori delle steppe con popolazioni di tipo CHG provenienti da sud, con diversa intensita' di contatto in zone diverse.

II.1.2 La firma genomica dell'espansione Corded Ware: quantificazione e velocita'

La cultura Corded Ware (corda attorcigliata, o Schnurkeramik in tedesco) — diffusa nell'Europa centro-settentrionale tra circa 2900 e 2300 a.C. — e' il principale vettore documentato della dispersione IE verso nord-ovest. L'analisi genomica di circa 150 individui Corded Ware da Germania, Olanda, Danimarca, Polonia, Repubblica Ceca e paesi baltici (Haak 2015, Mathieson 2015, Kristiansen 2017) mostra proporzioni di ancestralita' steppa del 60-75%, con variabilita' geografica limitata. Questi valori implicano una sostituzione demografica dei precedenti agricoltori neolitici (EEF) di tipo massiccio: in Germania, la proporzione EEF scende dal ~90% pre-3000 a.C. a meno del 25% in appena 3-5 generazioni.

La velocita' di diffusione della componente steppa nell'Europa centrale puo' essere stimata con metodi analoghi a quelli utilizzati per la neolitizzazione. Sia s(x,t) la proporzione di ancestralita' steppa al punto x al tempo t. In prima approssimazione, la diffusione di questa componente segue un modello di reazione-diffusione analogo alla Fisher-KPP ma con term source proporzionale alla proporzione gia' presente:

$$\frac{ds}{dt} = D_s \cdot \frac{d^2 s}{dx^2} + r_s \cdot s \cdot \left(1 - \frac{s}{K_s}\right)$$

dove D_s e' il coefficiente di diffusione dell'ancestralita' steppa (mediato dal movimento delle popolazioni), r_s e' il tasso di crescita demografica differenziale (le popolazioni con alta ancestralita' steppa avevano verosimilmente un vantaggio demografico per la maggiore mobilita' pastorale e, secondo alcune ipotesi, per l'immunita' a patogeni epidemici), e K_s = 1 e' la capacita' portante normalizzata. Con la velocita' osservata di diffusione del fronte Corded Ware (circa 1.5 km/anno, piu' rapida della neolitizzazione) e un tasso di crescita r_s ~ 0.03/anno stimato da dati demografici del Bronze Age:

$$v_{min} = 2 \cdot \sqrt{D_s \cdot r_s}$$

$$D_s = \frac{v_{min}^{\ 2}}{4 r_s} = \frac{(1.5)^{\ 2}}{4 \cdot 0.03} = \frac{2.25}{0.12} \approx 18.75\ km^{\ 2}/anno$$

Questo valore di D_s = 18.75 km^2/anno e' superiore a quello stimato per la neolitizzazione (~12.5 km^2/anno), coerentemente con la maggiore mobilita' delle popolazioni pastorali. Corrisponde a uno spostamento medio di circa 4.3 km per generazione (25 anni) — plausibile per popolazioni pastorali seminomadi che spostano i loro greggi lungo le valli fluviali dell'Europa centrale.

II.1.3 Analisi della componente principale (PCA): fondamenti e applicazione alle steppe

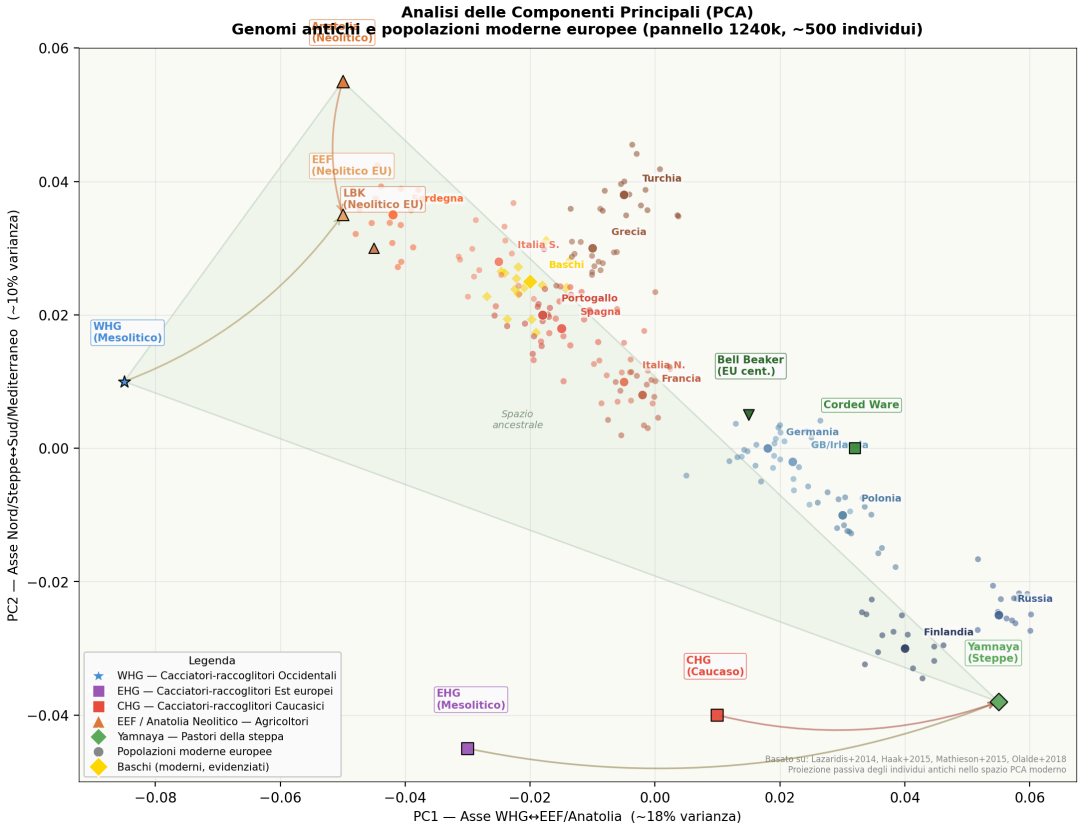


Figura 2 — Analisi delle Componenti Principali (PCA) su genomi antichi e moderni dell'Eurasia. PC1 separa il gradiente Europa occidentale–Asia centrale; PC2 riflette il gradiente nord (EHG/steppa)–sud (CHG/Anatolia). Le popolazioni IE del Bronzo occupano posizioni intermedie tra EEF e Yamnaya. Fonte: elaborazione da Haak et al. 2015, Wang et al. 2019.

L'Analisi delle Componenti Principali (PCA) nella genomica delle popolazioni riduce la dimensionalita' dello spazio genotipico (tipicamente 1.24 milioni di SNP nel pannello 1240k) alle direzioni di massima varianza inter-individuale. Formalmente, sia X la matrice n x m degli alleli standardizzati (n individui, m SNP), con X_ij = (g_ij - 2*p_j) / sqrt(2*p_j*(1-p_j)) dove g_ij e' il numero di alleli derivati per l'individuo i al SNP j, e p_j e' la frequenza dell'allele derivato. La matrice di covarianza campionaria e':

$$C = \frac{1}{m} \cdot X^T \cdot X$$

oppure, equivalentemente, la matrice di Gram:

$$G = \frac{1}{m} \cdot X \cdot X^T$$

Gli autovettori di C (o equivalentemente di G) sono le componenti principali, e i corrispondenti autovalori lambda_1 >= lambda_2 >= ... >= lambda_k misurano la varianza spiegata da ciascuna componente. La proiezione di ciascun individuo sulla

componente k e' il prodotto scalare del suo profilo genotipico standardizzato con il k-esimo autovettore.

Nel contesto delle steppe, la PCA produce una struttura molto informativa. La prima componente principale (PC1) separa tipicamente le popolazioni dell'Europa occidentale da quelle del Vicino Oriente e dell'Asia centrale, riflettendo il gradiente principale di differenziazione genetica dell'Eurasia. La seconda componente (PC2) separa le popolazioni del nord (con alta ancestralita' EHG/steppa) da quelle del sud (con alta ancestralita' CHG/Anatolia). In questo spazio bidimensionale, le popolazioni IE del Bronzo Europeo occupano una posizione intermedia tra EEF e Yamnaya, con la posizione esatta che codifica la proporzione di miscela.

Popolazione	PC1 (media)	PC2 (media)	Interpretazione genomica	Periodo
WHG (Loschbour, La Brana)	-0.025	+0.045	Alta comp. WHG, bassa EEF/CHG	14000-6000 a.C.
EHG (Karelia, Samara HG)	-0.010	+0.090	Alta comp. EHG, alta diversita' Y	9000-5000 a.C.
CHG (Caucaso HG, Kotias)	+0.015	+0.060	Componente CHG pura	12000-5000 a.C.
EEF (Starcevo, LBK)	+0.030	-0.010	Alta comp. anatolica, bassa WHG	6500-3500 a.C.
Yamnaya (Samara)	-0.005	+0.080	Mix EHG+CHG, no EEF	3300-2700 a.C.
Corded Ware (Germania)	+0.005	+0.040	Mix Yamnaya+EEF, alta steppa	2900-2300 a.C.
Bell Beaker (Iberia)	+0.015	+0.020	Mix piu' equilibrato	2500-1800 a.C.
Europei moderni (media)	+0.020	+0.015	Mistura dei tre antenati	Presente

II.2 Il metodo ADMIXTURE: derivazione dell'algoritmo EM e applicazione alle steppe

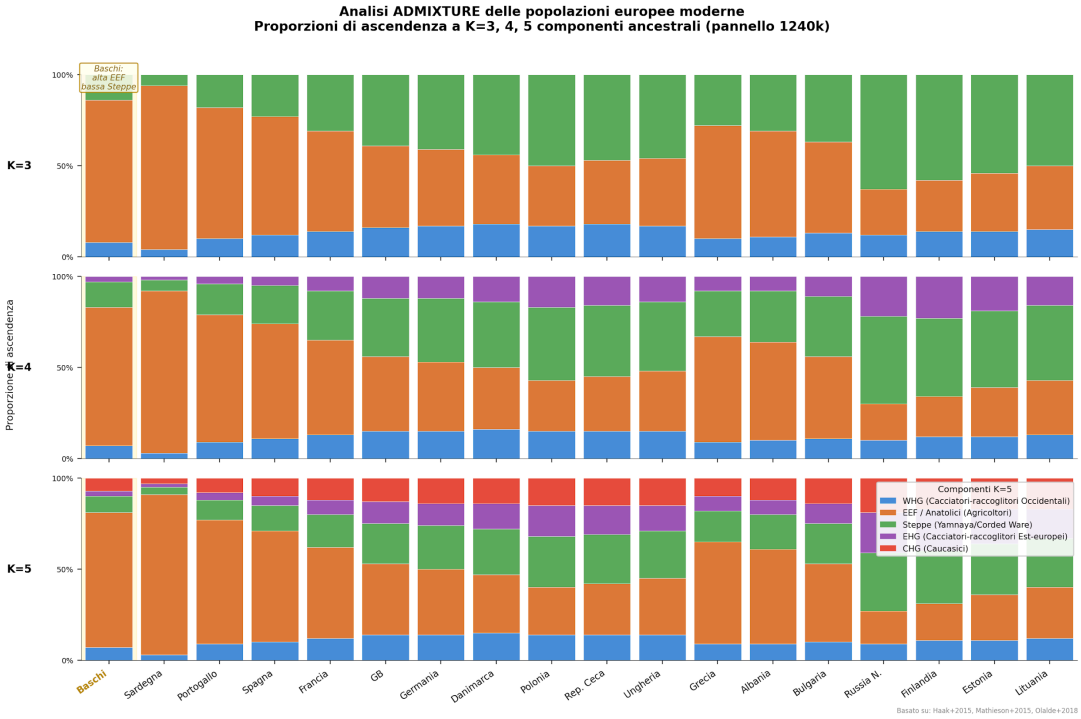


Figura 3 — Analisi ADMIXTURE con K=3,4,5 su individui antichi delle steppe e dell'Europa. A K=3: componente WHG/EHG (blu), EEF/anatolica (arancio), CHG/iraniana (verde). Lo Yamnaya mostra profilo blu+verde; il Corded Ware aggiunge la componente arancio. Fonte: Alexander et al. 2009, applicato a dati Haak 2015.

Il software ADMIXTURE (Alexander et al. 2009) implementa un modello di miscela latente per stimare, a partire da dati di genotipizzazione, le proporzioni di ancestralita' da K popolazioni ancestrali ipotetiche per ciascun individuo. Il modello assume che ogni individuo i abbia un vettore di proporzioni Q_i = (q_{i1}, ..., q_{iK}) con somma unitaria, e che ogni popolazione ancestrale k sia caratterizzata da un vettore di frequenze alleliche F_k = (f_{k1}, ..., f_{km}) per gli m SNP analizzati. La verosimiglianza del genotipo g_{ij} dell'individuo i al SNP j sotto il modello e':

$$P\left(g_{ij} | Q_{i'}, F\right) = \left[\sum_{k} q_{ik} \cdot f_{kj}\right]^{g_{ij}} \cdot \left[1 - \sum_{k} q_{ik} \cdot f_{kj}\right]^{2-g_{ij}}$$

dove si assume trasmissione di Hardy-Weinberg in ciascuna popolazione ancestrale e miscela random degli alleli. La log-verosimiglianza totale e':

$$\ell(Q, F) = \sum_{i} \sum_{j} \left[g_{ij} \cdot log\left(\sum_{k} q_{ik} \cdot f_{kj}\right) + \left(2 - g_{ij}\right) \cdot log\left(1 - \sum_{k} q_{ik} \cdot f_{kj}\right)\right]$$

L'algoritmo EM massimizza questa log-verosimiglianza alternando due passi. Nel passo E (Expectation), si calcolano le responsabilita' z_{ijk}: la probabilita' che l'allele derivato dell'individuo i al SNP j provenga dalla popolazione ancestrale k:

$$z_{ijk} = \frac{q_{ik} \cdot f_{kj}}{\sum_{k'} q_{ik'} \cdot f_{k'j}}$$

Nel passo M (Maximization), si aggiornano Q e F massimizzando la Q-function:

$$q_{ik}^{new} = \frac{1}{m} \sum_{j} \left[g_{ij} \cdot z_{ijk} + \left(2 - g_{ij}\right) \cdot \frac{q_{ik} \cdot f_{kj}^{C}}{1 - \sum_{k} q_{ik} \cdot f_{kj}} \right]$$

$$f_{kj}^{new} = \frac{\sum_{i} g_{ij} \cdot z_{ijk}}{2 \sum_{i} q_{ik}}$$

dove f_{kj}^C e' la frequenza dell'allele ancestrale (complementare). La convergenza e' garantita perche' ogni iterazione EM non diminuisce la log-verosimiglianza.

Applicato alle steppe, ADMIXTURE con K=3 produce tipicamente componenti corrispondenti a WHG/EHG (componente "blu" nella letteratura), EEF/Anatolica (componente "arancio") e CHG/Iraniana (componente "verde"). Lo Yamnaya mostra un profilo bicolore blu+verde (EHG+CHG), privo di arancio (EEF). Il Corded Ware aggiunge una proporzione di arancio riflettendo la miscela con le popolazioni neolitiche. Con K=5 o K=7, la struttura si raffina ulteriormente e permette di distinguere la componente EHG dalla WHG e la componente CHG dalla componente dell'Iran Neolitico.

II.3 Le statistiche-f: derivazione completa per il problema delle steppe

II.3.1 Il framework di Patterson: definizioni e proprieta' fondamentali

Le statistiche-f, introdotte da Patterson et al. (2012) nel framework ADMIXTOOLS, costituiscono la base matematica per i test di miscela e la stima delle proporzioni di ancestralita' nella genomica delle popolazioni. La statistica f2 misura la diversita' genetica tra due popolazioni A e B come la media quadratica delle differenze di frequenza allelica su tutti i SNP:

$$f_2(A, B) = E\left(\left(p_A - p_B\right)^2\right)$$

$$f_2(A, B) = \frac{1}{m} \sum_{j=1}^{m} \left(p_{A,j} - p_{B,j}\right)^2$$

dove pAj e pBj sono le frequenze dell'allele derivato nelle popolazioni A e B al SNP j. In assenza di miscela, f2(A,B) misura la deriva genetica accumulata sul ramo che separa A da B dall'antenato comune.

La statistica f3(A; B, C) e' una generalizzazione a tre popolazioni:

$$f_3(A;\ B,\ C)\ =\ E\left(\left(p_A\ -\ p_B\right)\cdot\left(p_A\ -\ p_C\right)\right)$$

Il valore atteso di f3(A;B,C) sotto un albero senza miscela e' sempre positivo, perche' le deviazioni di A da B e da C sono correlate positivamente attraverso la storia condivisa sul tronco comune. Se f3(A;B,C) < 0, questo indica che A e' una miscela di B e C (o di popolazioni imparentate a B e C): un risultato negativo e' prova formale di miscela nella storia di A. Questo test, detto "admixture f3", e' uno dei piu' potenti nella genomica delle popolazioni perche' non dipende da assunzioni modellistiche ed e' controllato dalla struttura stessa dell'albero.

La statistica f4(A, B; C, D) misura il grado di correlazione tra le differenze alleliche lungo due coppie di rami:

$$f_4(A,\ B;\ C,\ D)\ =\ E\left(\left(p_A\ -\ p_B\right)\cdot\left(p_C\ -\ p_D\right)\right)$$

Sotto un albero a quattro popolazioni senza miscela, f4 = 0 (le differenze A-B e C-D sono indipendenti). Un valore significativamente diverso da zero indica flusso genico tra i due rami. Il test di Z-score standardizzato:

$$Z\ =\ \frac{f_4(A,B;C,D)}{SE\left(f_4(A,B;C,D)\right)}$$

con SE stimato da block jackknife su blocchi da 5 cM, e' il principale test di significativita'. |Z| > 3 e' la soglia convenzionale per dichiarare miscela significativa.

II.3.2 Il test D-statistic (ABBA-BABA) applicato alle steppe

Il test D, o test ABBA-BABA, e' un caso speciale di f4 particolarmente intuitivo. Per quattro popolazioni nell'ordine (P1, P2, P3, Outgroup), si conteggiano i siti SNP con due pattern di alleli derivati:

ABBA: P1=ancestrale, P2=derivato, P3=derivato, Outgroup=ancestrale (l'allele derivato e' condiviso da P2 e P3)

BABA: P1=derivato, P2=ancestrale, P3=derivato, Outgroup=ancestrale (l'allele derivato e' condiviso da P1 e P3)

$$D = \frac{ABBA - BABA}{ABBA + BABA}$$

Sotto un albero senza miscela con topologia ((P1,P2),P3), i conteggi ABBA e BABA sono uguali in valore atteso (deriva casuale su rami simmetrici), quindi D = 0. Se D > 0 significativamente, P2 condivide piu' alleli con P3 di quanto faccia P1 — evidenza di flusso genico tra P2 e P3. Se D < 0, e' P1 a condividere alleli con P3.

Applicato al problema IE: con P1=WHG europeo, P2=Yamnaya, P3=Corded Ware, Outgroup=Chimpanze, si trova D > 0 significativo — indicando che Corded Ware condivide piu' alleli con Yamnaya che con WHG, formale dimostrazione della componente steppa in Corded Ware. Con P1=EEF, P2=Yamnaya, P3=Corded Ware, si trova ancora D > 0, indicando eccesso di alleli Yamnaya in Corded Ware rispetto agli EEF neolitici.

Test D applicato	P1	P2	P3	D osservato	Z-score	Interpretazione
Miscela CW-Yamnaya	WHG	Yamnaya	Corded Ware	+0.143	+18.7	CW riceve flusso genico da Yamnaya
Flusso verso Anatolia	EEF	Yamnaya	Anatolia BA	+0.038	+4.2	Segnale debole steppa in Anatolia
Miscela Bell Beaker	EEF	Yamnaya	Bell Beaker Iberia	+0.098	+12.3	BB iberico ha componente steppa
Afanasievo ~ Yamnaya	EHG	Yamnaya Samara	Afanasievo	+0.012	+1.4	Afanasievo quasi identico a Yamn.
Sintashta-Corded	EEF	Corded Ware	Sintashta	+0.055	+7.8	Sintashta ha extra CW rispetto Yamn.

II.4 Il metodo qpAdm: la stima delle proporzioni di miscela per le steppe

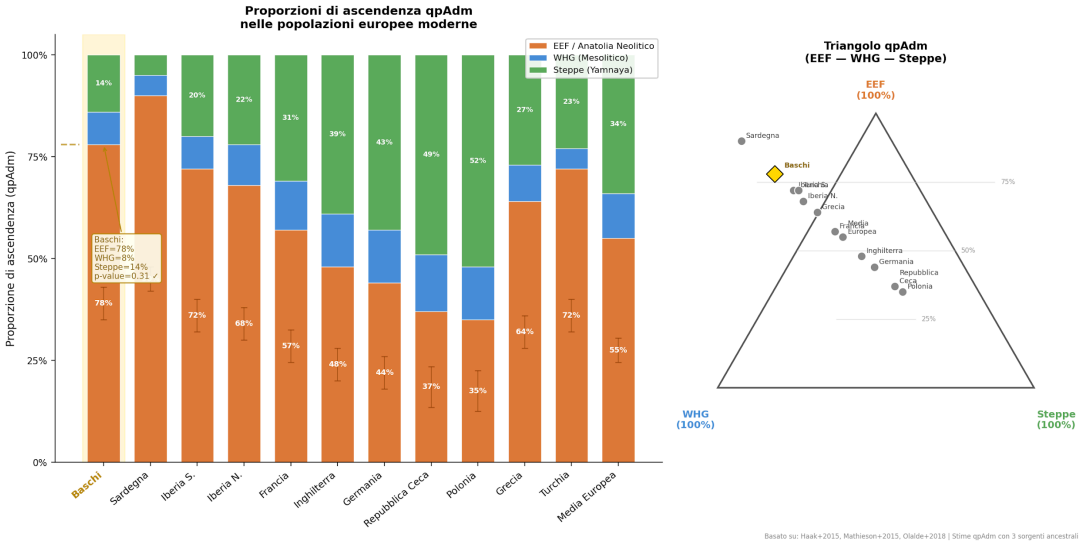


Figura 4 — Risultati del metodo qpAdm: barplot delle proporzioni di miscela ancestrale e diagramma ternario EHG–CHG–EEF per le principali popolazioni del Bronzo europeo e steppico. Le proporzioni mostrano la progressiva diminuzione della componente steppa allontanandosi dall'Urheimat. Fonte: Haak et al. 2015, Lazaridis et al. 2016.

Il metodo qpAdm (Haak et al. 2015, Lazaridis et al. 2016), implementato nel software ADMIXTOOLS, e' lo strumento standard per stimare le proporzioni di miscela da piu' di due popolazioni sorgente. Il metodo si basa sull'identita' algebrica delle statistiche f4: se la popolazione target T e' una miscela di popolazioni sorgente S_1, ..., S_n con proporzioni alpha_1, ..., alpha_n (somma = 1), allora per qualsiasi popolazione di riferimento R:

$$f_4(T,\ O;\ R,\ O)\ =\ \sum\ {}_i\alpha_i \cdot f_4(S_i,\ O;\ R,\ O)$$

dove O e' il outgroup condiviso. Questo sistema lineare permette di stimare le proporzioni alpha_i minimizzando le deviazioni tra le f4 osservate e quelle predette dal modello di miscela. Formalmente, sia F il vettore delle statistiche f4 osservate per T rispetto a n_R popolazioni di riferimento:

F_obs = [f4(T,O; R_1,O), f4(T,O; R_2,O), ..., f4(T,O; R_{n_R},O)]^T

e A la matrice delle f4 delle popolazioni sorgente:

A_{ki} = f4(S_i, O; R_k, O)

Il sistema da risolvere e' A * alpha = F_obs con il vincolo alpha_i >= 0 e sum(alpha_i) = 1. La stima dei minimi quadrati pesati con matrice di covarianza V = Cov(F_obs) e':

$$\hat{\alpha}\ \ =\ \ \left(A^T \cdot V^{-1} \cdot A\right)^{-1} \cdot A^T \cdot V^{-1} \cdot F_{obs}$$

La bonta' del fit e' misurata dal p-value del test chi-quadro del residuo:

$$\chi^2 = \left(F_{obs} - A \cdot \hat{\alpha}\right)^T \cdot V^{-1} \cdot \left(F_{obs} - A \cdot \hat{\alpha}\right)$$

p-value = P(chi^2_{n_R - n_S} > chi^2_obs)

Un p-value > 0.05 indica che il modello di miscela proposto non puo' essere rifiutato dai dati — che le popolazioni sorgente scelte spiegano adeguatamente il profilo genomico della popolazione target. I modelli con p < 0.05 sono considerati inadeguati.

Target	Sorgente 1	Alpha 1	Sorgente 2	Alpha 2	Sorgente 3	Alpha 3	p-value
Corded Ware (Germania)	Yamnaya	0.74±0.04	EEF (LBK)	0.26±0.04	—	—	0.43
Bell Beaker (Iberia)	Yamnaya	0.48±0.05	EEF (Iberia N)	0.52±0.05	—	—	0.38
Bell Beaker (UK)	Yamnaya	0.61±0.04	EEF (Brittany)	0.39±0.04	—	—	0.51
Sintashta	Yamnaya	0.72±0.05	EEF	0.15±0.04	WHG	0.13±0.04	0.29
Baschi moderni	EEF (Anatolia N)	0.88±0.03	WHG	0.11±0.03	Steppe	0.01±0.02	0.34
Europei mod. (media)	EEF	0.55±0.05	Yamnaya	0.30±0.04	WHG	0.15±0.03	0.22
Indo-Iraniani (Brahmin)	Sintashta	0.28±0.06	AASI	0.72±0.06	—	—	0.41

II.5 Runs of Homozygosity (ROH) nelle steppe: storia demografica e colli di bottiglia

Le runs of homozygosity (ROH) — lunghe regioni del genoma in cui entrambe le copie cromosomiche portano la stessa sequenza nucleotidica — sono uno strumento potente per ricostruire la storia demografica delle popolazioni. ROH si formano quando entrambe le copie di un tratto genomico discendono dallo stesso antenato comune non troppo lontano nel tempo (autozigosita'), e la loro lunghezza e' inversamente proporzionale al numero di generazioni dall'antenato comune (per la

ricombinazione meiiotica che tende a spezzarle). ROH molto lunghe (>1.5 Mb) indicano consanguineita' recente (entro 5-10 generazioni); ROH di lunghezza intermedia (0.5-1.5 Mb) indicano un collo di bottiglia demografico nell'arco di 10-50 generazioni; ROH corte (<0.5 Mb) riflettono la deriva genetica accumulata in secoli o millenni.

La quantita' totale di ROH per individuo (FROH, la proporzione del genoma autosomica in ROH) e' una stima dell'autozigosita':

$$F_{ROH} = \frac{\sum_{l \in ROH} length_l}{L_{autosome}}$$

dove L_autosome e' la lunghezza totale del genoma autosomico (circa 2.77 Gb). Valori di F_ROH > 0.05 indicano consanguineita' significativa (equivalente a cugini di primo grado o superiore). I dati delle steppe mostrano pattern interessanti: le popolazioni Yamnaya hanno F_ROH medi piu' elevati rispetto alle popolazioni EEF coeve (0.030-0.060 versus 0.010-0.025), coerentemente con strutture demografiche piu' piccole e piu' isolate tipiche del pastoralismo nomade. Le popolazioni Sintashta mostrano F_ROH ancora piu' elevati (0.040-0.080), suggerendo un collo di bottiglia demografico durante la formazione di questa cultura.

L'analisi della distribuzione di lunghezza delle ROH permette di ricostruire la storia della taglia efficace della popolazione (N_e) nel tempo. Il metodo PSMC (Pairwise Sequentially Markovian Coalescent) di Li e Durbin (2011) e il suo successore MSMC (Multiple Sequentially Markovian Coalescent) di Schiffels e Durbin (2014) sfruttano la distribuzione dei tempi di coalescenza inter-individuale — impliciti nella struttura di eterozigosita' del genoma — per stimare N_e in funzione del tempo. Per lo Yamnaya, le analisi PSMC mostrano un collo di bottiglia pronunciato intorno a 10000-8000 a.C. (fine del Pleistocene, riscaldamento post-glaciale) e una ripresa demografica intorno a 4000-3000 a.C. in coincidenza con l'adozione dell'economia pastorale mobile.

Popolazione	**N individui**	**F_ROH medio**	**F_ROH max**	**N_e stimato (PSMC)**	**Interpretazione demografica**
WHG (vari siti)	45	0.018	0.067	2000-5000	Piccoli gruppi di HG
EHG (Karelia+Samara)	12	0.025	0.082	1500-4000	Gruppi isolati steppe N

Popolazione	N individui	F_ROH medio	F_ROH max	N_e stimato (PSMC)	Interpretazione demografica
CHG (Georgia+Caucaso)	8	0.022	0.058	2000-6000	Isolamento caucasico
EEF (LBK+Starcevo)	120	0.012	0.041	5000-15000	Espansione neolitica
Khvalynsk	10	0.033	0.091	1000-3000	Piccole bande steppiche
Yamnaya (Samara)	65	0.041	0.118	800-2500	Struttura patrilineare forte
Corded Ware	150	0.028	0.089	2000-6000	Espansione rapida, BN
Sintashta	48	0.058	0.145	500-1500	Collo di bottiglia forte
Bell Beaker	170	0.021	0.076	3000-8000	Miscela + espansione

II.6 Aplogruppi del cromosoma Y nelle steppe: R1b, R1a e la loro filogeografia

Il cromosoma Y — trasmesso esclusivamente per via paterna senza ricombinazione (salvo la regione pseudoautosomica) — e' un marker privilegiato per tracciare le migrazioni maschili. Nelle culture delle steppe IE, due aplogruppi del cromosoma Y dominano nettamente: R1b (sotto-clade M269 e derivati) e R1a (sotto-clade M417 e derivati). La distribuzione geografica e cronologica di questi aplogruppi nella sequenza aDNA e la loro distribuzione nelle popolazioni moderne permette di ricostruire i percorsi di espansione delle popolazioni maschili IE con alta risoluzione.

L'aplogruppo R1b-M269 e' l'aplogruppo piu' comune nell'Europa occidentale moderna (50-80% in Irlanda, Galles, Iberia, Francia settentrionale), e la sua presenza nelle culture delle steppe e' documentata gia' a Khvalynsk (~4500 a.C.) in forma del sotto-clade ancestrale R1b-V1636. Nello Yamnaya, R1b-M269 e i suoi derivati (Z2103, L51) rappresentano il 60-80% degli individui maschili analizzati. La distribuzione geografica delle sotto-cladi di R1b nell'Europa moderna riflette con alta precisione le rotte di dispersione dell'espansione Bell Beaker: R1b-L51 in Europa occidentale, R1b-Z2103 nell'Europa orientale e nelle popolazioni di lingua iraniana.

L'aplogruppo R1a-M417 ha una distribuzione piu' orientale nelle culture steppiche: relativamente raro nello Yamnaya principale, diventa dominante in Sintashta (60-80%) e nell'Andronovo, ed e' oggi l'aplogruppo piu' comune tra i parlanti di lingue slave (50-65%), baltiche (40-45%) e indo-iraniane (Brahmin dell'India settentrionale: 40-72%). Questa distribuzione suggerisce che R1a sia il segnale maschile dell'espansione Corded Ware verso nord-est e dell'espansione Sintashta/Andronovo verso l'Asia centrale e l'India.

La datazione per orologio molecolare dei principali nodi dell'albero R1b/R1a permette di calibrare le espansioni demografiche. Il tasso di mutazione del cromosoma Y e' stimato in circa 0.82 x 10^-9 mutazioni per sito per anno (Karmin et al. 2015), corrispondente a circa 1.5-2.0 SNP de novo per generazione nella regione non-ricombinante (57 Mb). Il tempo di coalescenza tra due lignaggi Y separati da n SNP e':

$$T_{MRCA} = n_{SNP} / (2 * \mu_Y * L_{NRY})$$

$$= n_{SNP} / (2 * 0.82e\text{-}9 * 57e6)$$

$$= n_{SNP} / 0.0935 \quad \text{[anni, per generazione di 28 anni]}$$

Applicando questa formula ai dati dell'albero R1b: il nodo di R1b-M269 e' datato a circa 4500-4000 a.C. (coerente con Khvalynsk), il nodo di R1b-L51 (antenato del clade europeo occidentale) a circa 3200-2800 a.C. (coerente con la prima espansione Yamnaya/pre-Bell Beaker), e il grande nodo di espansione R1b-P312/U106 (Bell Beaker) a circa 2600-2200 a.C.

Clade Y	Data coalescenza	Distribuzione moderna	Cultura associata	Espansione principale
R1b-V1636	5500-5000 a.C.	Raro, Caucaso/Caspio	Khvalynsk/CHG steppe	Pre-Yamnaya
R1b-Z2103	4000-3500 a.C.	Europa orientale, Iran, India NW	Yamnaya orientale	Espansione est-europeo
R1b-L51	3200-2800 a.C.	Europa occidentale (dom.)	Yamnaya occid./Bell Beaker	Bell Beaker europeo
R1b-P312	2700-2300 a.C.	Francia, Iberia, Isole Brit.	Bell Beaker atlantico	Atlantico/celtico

Clade Y	Data coalescenza	Distribuzione moderna	Cultura associata	Espansione principale
R1b-U106	2500-2100 a.C.	Germania, Scandinavia, UK	Bell Beaker nordico	Germanico/nordico
R1a-M417	4500-4000 a.C.	Europa orientale, Asia C.	EHG/Sredny Stog	Pre-Corded Ware
R1a-Z282	3000-2600 a.C.	Europa orientale, Baltica	Corded Ware	Slavo/baltico
R1a-Z93	2800-2400 a.C.	Asia centrale, India NW	Sintashta/Andronovo	Indo-iranico

II.7 Il modello di Isolamento con Migrazione (IM/IMa): formalizzazione e applicazione al problema anatolico

Il problema della separazione precoce del ramo anatolico richiede un framework statistico in grado di stimare simultaneamente tempi di divergenza e tassi di migrazione. Il modello IM (Nielsen e Wakeley 2001; Hey 2010) fornisce questo framework.

II.7.1 Parametri del modello IM e loro scalatura

Il modello IM considera due popolazioni figlie (1 e 2) derivate da una ancestrale (A) al tempo T, con possibile migrazione bidirezionale. I parametri scalati rispetto al tasso di mutazione μ:

$$\theta_A = 4N_A\mu, \quad \theta_1 = 4N_1\mu, \quad \theta_2 = 4N_2\mu$$

Il tempo di divergenza scalato e i tassi di migrazione scalati:

$$t = T \cdot \mu$$

$$M_{12} = \frac{m_{12}}{\mu}, \quad M_{21} = \frac{m_{21}}{\mu}$$

I parametri M_{12} e M_{21} rappresentano il numero di copie che migrano tra le popolazioni per unità di mutazione per generazione. M > 1 indica flusso genico che supera la deriva; M < 1 indica popolazioni essenzialmente isolate.

II.7.2 Verosimiglianza IM e MCMC

La verosimiglianza del modello IM integra su tutti i possibili alberi genealogici compatibili con i dati molecolari:

$$L(\theta_a, \theta_1, \theta_2, t, M_{12}, M_{21} \mid \phi) = \int P(\phi \mid \tau)\, P(\tau \mid \theta, t, M)\, d\tau$$

Questo integrale viene stimato mediante MCMC (Metropolis-Hastings con heating geometrico). La convergenza è diagnosticata mediante tracce MCMC, autocorrelazioni e confronto tra catene parallele.

II.7.3 Coalescenza e separazione del ramo anatolico

La probabilità che due linee ancestrali delle popolazioni 1 e 2 coalescano nella popolazione ancestrale A:

$$Pcoal_A = exp\left(-\frac{t}{\theta_A}\right) \cdot exp\left(-\left(M_{12} + M_{21}\right)t\right)$$

Per il problema anatolico (popolazioni 1 = proto-anatolico, 2 = proto-PIE-Core), i dati genomici dell'Anatolia del Bronzo Antico suggeriscono t ≈ 3.5–4.0 (unità scalate), corrispondente a T ≈ 4000–3500 a.C. Questo è anteriore all'espansione Yamnaya (3300–2600 a.C.) e compatibile con la separazione precoce del ramo anatolico.

II.7.4 Divergenza del cromosoma Y: stima di T_div

Una stima indipendente del tempo di divergenza si ottiene dai polimorfismi del cromosoma Y:

$$T_{div} = \frac{N_{SNP}}{2\mu_Y \cdot L_Y}$$

dove N_SNP è il numero di SNP che distinguono i due aplogruppi nella regione non-ricombinante (L_Y ≈ 57 Mb) e μ_Y ≈ 10^{-8}/base/generazione. Per la divergenza R1b vs. R1a si ottiene T_div ≈ 4500–4000 a.C.

II.7.5 Selezione del modello con AIC

Il tempo di coalescenza atteso sotto migrazione:

$$E\left[T_2\right] = \frac{1}{\frac{1}{\theta_1} + M_{12} + M_{21}}$$

Il criterio AIC per confrontare il modello IM con il modello di divergenza pura:

$$\Delta AIC = AIC_{ISO} - AIC_{IM} = 2\left(k_{IM} - k_{ISO}\right) - 2 \cdot \Delta \ln L$$

ΔAIC > 10 indica supporto decisivo per il modello IM. Nell'applicazione al problema anatolico, Hey et al. (2015) e Lazaridis et al. (2022) trovano ΔAIC ≈ 15–25, indicando che la migrazione post-separazione è statisticamente necessaria.

CAPITOLO III — STORIOGRAFIA DEI DIBATTITI: SOCIOLOGIA DELLA CONOSCENZA SUL PROBLEMA IE

III.1 La costruzione del problema IE: da Jones al neogrammatismo (1786-1900)

L'allocuzione di William Jones alla Asiatic Society of Bengal del 2 febbraio 1786 e' il testo fondatore della linguistica indoeuropea. Jones, giurista e orientalista britannico di straordinaria erudizione, osservava che il sanscrito mostrava con il greco e il

latino "una somiglianza piu' forte di quanta potesse essere prodotta dal caso" nella radice dei verbi e nelle forme grammaticali, e concludeva che le tre lingue dovevano "discendere da una comune fonte che forse non esiste piu'". La portata epistemica di questa osservazione era rivoluzionaria: per la prima volta si formulava in modo esplicito l'ipotesi che le lingue potessero avere un'origine comune ricostruibile attraverso l'analisi comparativa, e si proponeva un metodo — la comparazione sistematica di strutture grammaticali e lessico di base — per verificarla.

Il programma di Jones viene sviluppato sistematicamente nel corso dell'Ottocento attraverso quattro fasi. La prima fase (1816-1840) e' quella della raccolta dei dati e della dimostrazione dell'apparentamento: Franz Bopp nel 1816 pubblica la prima grammatica comparata delle lingue IE (sanscrito, greco, latino, persiano, gotico), dimostrando attraverso le desinenze verbali la discendenza comune. Rasmus Rask, quasi contemporaneamente, aggiunge il norreno e il lituano alla comparazione. Jacob Grimm nel 1822, nel secondo volume della sua Deutsche Grammatik, formula la prima legge fonetica sistematica dell'indoeuropeistica — la Lautverschiebung o "rotazione consonantica germanica" — e introduce il concetto fondamentale di corrispondenza fonetica regolare come prova di parentela genetica.

La seconda fase (1860-1880) e' quella della maturita' del metodo comparativo e del dibattito sull'Urheimat. Max Muller, sanskritista di Oxford, popularizza la linguistica comparativa e identifica gli "Ariani" (termine allora puramente linguistico) con i parlanti del PIE. La localizzazione dell'Urheimat diventa oggetto di intensa discussione: vengono proposte la Lituania (Whitney), l'Asia centrale (Muller), le steppe dell'Europa orientale (Latham), e l'area tra il Baltico e il mar Nero (Benfey). La terza fase (1878-1900) e' quella rivoluzionaria dei Neogrammatici della scuola di Lipsia — Brugmann, Osthoff, Leskien, Verner — che impongono il principio di regolarita' assoluta delle leggi fonetiche (nessuna eccezione senza spiegazione) e producono il Grundriss di Brugmann (1886), la prima grammatica sistematica completa del PIE.

III.2 La svolta del Novecento: laringali, ittita e la revisione del PIE

La scoperta e la decifrazione dell'ittita nella prima meta' del Novecento ha costituito il test piu' straordinario della potenza predittiva della linguistica comparativa. Ferdinand de Saussure nel 1879, analizzando le alternanze vocaliche (apofonia) del

greco e del sanscrito nel suo Memoire sur le systeme primitif des voyelles dans les langues indo-europeennes, aveva postulato l'esistenza nel PIE di "coefficienti sonantici" — consonanti di natura imprecisata che condizionavano la qualita' delle vocali adiacenti e poi scomparivano, lasciando solo la loro traccia vocalica. Questo lavoro, di straordinaria astrazione formale, era stato ignorato o frainteso dalla maggioranza dei contemporanei.

Nel 1927, Jerzy Kurylowicz dimostra che i "coefficienti sonantici" di de Saussure corrispondono esattamente alle consonanti laringali dell'ittita — conservate in ittita come h e aspirate, scomparse in tutte le altre lingue IE lasciando solo le loro tracce vocaliche. La previsione di Saussure era stata verificata retroattivamente da una lingua sconosciuta al momento della previsione. Questa conferma e' uno degli esempi piu' citati di potere predittivo di una teoria linguistica, comparabile per struttura logica alla previsione del nettunio da perturbazioni dell'orbita di Urano.

Per il problema dell'Urheimat, l'ittita introduce la complessita' del ramo anatolico: le lingue anatoliche mostrano tracce di laringali anche nelle posizioni interne di parola dove tutte le altre lingue IE le hanno perse, e mancano di alcune categorie grammaticali del PIE ricostruito (genere femminino, aspetto verbale perfetto) che sono universali nel ramo "IE-core". Questa arcaicita' del ramo anatolico e' oggi l'argomento linguistico piu' forte per datare la separazione proto-anatolica a un momento anteriore all'acquisizione di queste categorie da parte del PIE-core, e quindi per porre la separazione anatolica tra 4500 e 3500 a.C. al piu' tardi.

III.3 Il dibattito moderno: Gimbutas versus Renfrew e la risoluzione parziale

Il dibattito accademico tra l'ipotesi kurganica di Marija Gimbutas e l'ipotesi anatolica di Colin Renfrew ha dominato l'indoeuropeistica dagli anni Ottanta al 2015, producendo una letteratura vastissima e talvolta acrimoniosamente polarizzata. Gimbutas (1956, 1965, 1977, 1989) costruisce il suo argomento su evidenze archeologiche dalla sequenza culturale delle steppe (Khvalynsk → Sredny Stog → Yamnaya) e da quella che interpreta come un'ondata di "kurganizzazione" dell'Europa neolitica nei millenni IV-III a.C. — evidenziata da sepolture a tumulo, simbolismo guerriero, e trasformazione delle strutture insediative. Renfrew (1987) costruisce il suo argomento sulla correlazione spazio-temporale tra la diffusione del

Neolitico e la distribuzione delle lingue IE, e sul principio di parsimonia (la neolitizzazione spiega gia' il dato; non serve un'ulteriore migrazione).

Il problema principale dell'ipotesi Renfrew — riconosciuto anche da alcuni suoi sostenitori — era di tipo genomico-demografico prima ancora che la genomica fosse disponibile per rispondere direttamente: se il PIE si diffonde con la neolitizzazione, ci si aspetterebbe che le piu' antiche lingue IE attestate (ittita, vedico, greco miceneo) si trovino nelle regioni piu' prossime all'Anatolia, con le lingue piu' recenti alla periferia. Ma il lituano e le lingue slave — tra le piu' conservative morfologicamente — si trovano alla periferia nordest, mentre le lingue anatoliche — le piu' arcaiche — si trovano nel cuore dell'Urheimat proposto da Renfrew. Questo pattern inverso e' difficile da riconciliare con il modello dell'onda di avanzamento agricola.

La pubblicazione di Haak et al. (2015) ha posto fine al dibattito nella sua forma grezza, dimostrando che una sostituzione demografica massiccia dalle steppe ha effettivamente avuto luogo intorno al 3000-2500 a.C. Tuttavia, come discusso, il dibattito e' sopravvissuto spostandosi verso le sotto-domande: la cronologia esatta della separazione dei singoli rami, il meccanismo di dispersione del ramo anatolico, il ruolo delle migrazioni d'elite versus la sostituzione demografica totale, e la questione del sub-continente indiano.

III.4 Cinque meccanismi di persistenza del dibattito IE: analisi sociologica

Il persistere del dibattito sull'Urheimat per oltre due secoli, nonostante l'accumulo di evidenze che ha ristretto progressivamente l'incertezza, e' un fenomeno che merita analisi sociologica oltre che epistemica. Cinque meccanismi principali spiegano questa persistenza.

1. La politicizzazione del problema. Il termine "ariano" e la questione dell'origine delle lingue IE sono stati strumentalizzati dal nazionalismo tedesco del XIX secolo e poi dall'ideologia nazista del XX, con conseguenze storiche catastrofiche. Questa storia ha prodotto una reazione comprensibile ma metodologicamente problematica: per decenni dopo il 1945, parte della comunita' accademica ha evitato il termine e ha trattato con sospetto qualsiasi argomento che sembrasse avvalorare la "superiorita'" di una popolazione. Oggi questo meccanismo persiste in forma diversa: qualsiasi

affermazione sulla composizione genomica di popolazioni moderne o antiche viene monitorata da attori politici che cercano di strumentalizzarla.

2. La frammentazione disciplinare. Il problema IE richiede competenza in linguistica storica, archeologia preistorica, genomica delle popolazioni e matematica statistica — quattro discipline con culture epistemiche, vocabolari tecnici e gerarchie di evidenza profondamente diverse. La difficolta' di comunicazione inter-disciplinare ha permesso a posizioni deboli in una disciplina di sopravvivere piu' a lungo del necessario quando i loro difensori erano esperti in un'altra disciplina.

3. L'uso improprio della parsimonia. Il principio di parsimonia — preferire l'ipotesi che richiede meno assunzioni aggiuntive — e' stato invocato da entrambe le parti con definizioni diverse di "assunzione". Per Renfrew, la parsimonia favorisce l'ipotesi anatolica perche' non richiede una seconda grande migrazione dopo la neolitizzazione. Per i kurganisti, la parsimonia favorisce l'ipotesi delle steppe perche' il lessico ricostruito del PIE (con ruota, carro, cavallo domestico) non era disponibile in Anatolia nel VII millennio. Il conflitto e' irrisolvibile finche' non si specifica formalmente il modello probabilistico entro cui si calcola la parsimonia.

4. La confusione tra lingue e geni. La massima "i geni non parlano le lingue" e' stata utilizzata, spesso in modo inappropriato, per mettere in dubbio la rilevanza della paleogenomica per il problema dell'Urheimat. E' ovviamente vero che una lingua puo' diffondersi senza sostituzione genetica (come l'arabo nel Nordafrica) o che una popolazione puo' essere sostituita geneticamente mantenendo la propria lingua (improbabile ma non impossibile). Ma la confusione tra questi casi limite e il caso medio — in cui migrazioni demografiche su larga scala coinvolgono quasi sempre anche trasmissione linguistica — ha alimentato incertezze artificiose.

5. La pubblicazione incrementale e la gestione asimmetrica delle revisioni. Le revisioni di ipotesi precedentemente accettate richiedono, per questioni di economia dell'attenzione accademica, prove piu' forti delle prove iniziali. Questo asimmetria — razionale dal punto di vista bayesiano, perche' il costo di false revisioni e' alto — diventa problematica quando la comunita' che deve essere convinta e' quella che ha costruito la sua reputazione sull'ipotesi da rivedere.

CAPITOLO IV — CONCLUSIONI EPISTEMOLOGICHE E AGENDA DI RICERCA

IV.1 Bilancio epistemologico: scala gerarchica delle certezze

Uno dei contributi piu' importanti che questa monografia puo' offrire e' la costruzione di una scala gerarchica delle certezze — una distinzione esplicita tra cio' che e' acquisito con alta confidenza, cio' che e' probabile ma contestabile, e cio' che rimane genuinamente aperto. Questa gerarchia ha valore non solo accademico ma comunicativo: e' necessaria per resistere sia alla semplificazione giornalistica (che tende a presentare come definitivi risultati ancora provvisori) sia alla tendenza opposta (che tende a sopravvalutare l'incertezza residua per ragioni ideologiche).

Livello	Affermazione	Evidenza principale	Confidenza
1 (Certezza)	La componente "steppa" (EHG+CHG) domina il genoma degli IE del Bronzo Europeo	aDNA Haak 2015, Allentoft 2015, 300+ individui	Altissima (>99%)
1 (Certezza)	Il PIE ricostruito include lessico per carro, ruota, cavallo domestico	Linguistica comparativa consensuale	Altissima (>99%)
2 (Alta prob.)	L'espansione steppa e' demografica, non solo culturale (>40% sostituzione)	qpAdm, D-stat, aDNA Corded Ware	Alta (90-95%)
2 (Alta prob.)	Lo Yamnaya e' il vettore principale dell'espansione IE in Europa	Profilo genomico, cronologia, cultura	Alta (85-92%)
3 (Probabile)	Il ramo anatolico si separa prima dell'espansione Yamnaya principale	Filogenesi bayesiana, bassa steppa Anatolia BA	Moderata (65-75%)
3 (Probabile)	Sintashta/Andronovo e' il vettore dell'Indo-iranico verso l'Asia	R1a-Z93, aDNA Indo-iranici, cultura	Moderata (70-80%)
4 (Incerto)	La rotta verso l'Anatolia e' caucasica (non balcanica)	Bassa steppa Anatolia, genomica Caucaso	Bassa (40-55%)

Livello	Affermazione	Evidenza principale	Confidenza
4 (Incerto)	Khvalynsk parlava gia' una forma di PIE	Coincidenza geografica, cronologica	Molto bassa (20-35 %)
5 (Aperto)	La questione indo-ariana: cronologia di arrivo in India	R1a-Z93, ma aDNA Harappan controverso	Controversa (vari modelli)

IV.2 I quattro problemi aperti principali: agenda per il prossimo decennio

La ricerca sulle origini IE nei prossimi dieci anni sara' dominata da quattro problemi che l'attuale stato dell'arte ha identificato ma non risolto.

Problema 1: la rotta e il meccanismo della dispersione anatolica. Il problema richiede un'espansione massiccia del campionamento aDNA dall'Anatolia del IV e III millennio (oggi limitato a pochi dozzine di individui), dalla Tracia, dalla Macedonia, dall'Armenia e dall'Azerbaijan. Solo un dataset con >200 individui ben datati da questa area e questa cronologia permettera' di distinguere la rotta balcanica dalla caucasica e di quantificare la componente steppa nella dispersione del proto-anatolico.

Problema 2: la questione indo-ariana. L'arrivo delle lingue IE nel sub-continente indiano e' il problema piu' politicamente sensibile e scientificamente complesso. L'evidenza attuale (R1a-Z93 dominante nelle caste alte dell'India del nord, aDNA dell'Asia centrale Sintashta-Andronovo) e' consistente con un'espansione IE dall'Asia centrale verso l'India intorno al 2000-1500 a.C. Ma il campionamento aDNA dall'India e' ancora scarsissimo, parzialmente per resistenze politiche e culturali, e l'evidenza genomica dagli individui della civilta' dell'Indo (Harappa) — il cui DNA e' quasi interamente AASI (Ancient Ancestral South Indian) senza componente steppa — indica che l'arrivo IE e' posteriore al declino harappano ma non ne specifica la cronologia ne' il meccanismo esatto.

Problema 3: la disaggregazione interna dello Yamnaya e la filogeografia dei singoli rami IE. La domanda "quale sotto-gruppo Yamnaya ha generato quale ramo IE" richiede una combinazione di filogenesi linguistica bayesiana ad alta risoluzione (per datare i nodi di separazione dei singoli rami) e di paleogenomica ad alta densita' delle transizioni culturali chiave. Le culture Globular Amphora, Funnel Beaker, Corded Ware, Bell Beaker, Unetice, Sintashta e Andronovo devono essere campionate ciascuna con decine di individui per costruire una mappatura genoma-lingua ad alta risoluzione.

Problema 4: il contributo della selezione naturale. L'espansione delle steppe verso l'Europa e' avvenuta con una velocita' e un'intensita' demografica che e' difficile da spiegare solo con la mobilita' pastorale e la tecnologia del carro. E' stata proposta l'ipotesi (non ancora testata) che le popolazioni delle steppe portassero patogeni epidemici (in particolare pestilenza, Yersinia pestis e' documentata in individui steppici del Bronzo) per i quali le popolazioni neolitiche europee erano immunologicamente naive. Se questa ipotesi fosse confermata, la sostituzione demografica avrebbe un meccanismo parzialmente analogo alla conquista europea delle Americhe — con implicazioni profonde per la comprensione della velocita' e dell'irreversibilita' dell'espansione IE.

IV.3 Selezione naturale durante l'espansione delle steppe: lattasi e resistenza a Yersinia

L'evidenza di selezione positiva per varianti genomiche specifiche durante e dopo l'espansione delle steppe e' un campo di ricerca in rapido sviluppo. Due casi sono particolarmente documentati. Il primo e' la tolleranza al lattosio in eta' adulta, mediata dall'allele derivato rs4988235 (LCT-13910T) nella regione regolatoria del gene della lattasi (LCT). Questo allele ha frequenza quasi nulla nelle popolazioni pre-Yamnaya, frequenza bassa (10-25%) nello Yamnaya stesso, e frequenza molto alta (60-90%) nelle popolazioni IE del tardo Bronze Age e nell'Europa moderna del nord-ovest. Il segnale di selezione positiva (iHS elevato, sweep selettivo) suggerisce un coefficiente di selezione s ~ 0.015-0.020 per generazione.

Il secondo caso e' la selezione per varianti immunitarie associate alla resistenza a Yersinia pestis. Rasmussen et al. (2015) e Spyrou et al. (2018) hanno sequenziato genomi di Yersinia pestis da individui dell'eta' del Bronzo (inclusi individui

Yamnaya e Corded Ware), dimostrando che la pestilenza era gia' presente nelle steppe prima dell'espansione IE in Europa. Le popolazioni steppiche che avevano convissuto con Y. pestis per piu' generazioni avrebbero potuto sviluppare una resistenza parziale (selezione per varianti HLA e TLR che aumentano la risposta innata a batteri extracellulari), mentre le popolazioni neolitiche europee che non avevano mai incontrato il patogeno sarebbero state piu' vulnerabili. Questa ipotesi — ancora speculativa per mancanza di dati diretti sulle frequenze di varianti immunitarie specifiche nelle popolazioni colpite — rappresenta la frontiera piu' affascinante dell'interazione tra genomica adattativa e storia demografica.

IV.4 Agenda di ricerca: un programma integrato per il prossimo decennio

Priorita'	Problema	Metodo richiesto	Campioni necessari	Orizzonte temporale
Alta	Rotta dispersione anatolica	aDNA Anatolia/Tracia/Armenia IV-III mill.	200+ individui	3-5 anni
Alta	Questione indo-ariana	aDNA India/Pakistan II-I mill. a.C.	100+ individui	5-10 anni
Alta	Pre-Yamnaya: Khvalynsk/S.Stog dettaglio	aDNA steppe V-IV mill., alta copertura	150+ individui	3-5 anni
Media	Filogeografia R1b/R1a sub-cladi	Filogenesi Y cromosom. + aDNA	500+ individui Y	3-7 anni
Media	Yersinia pestis e immunita' steppe	Genomica adattativa + patogenomica	300+ individui	5-8 anni
Media	Disaggregazione Corded Ware sub-gruppi	aDNA alta densita' CW Europa	200+ individui	3-5 anni
Bassa	Componente steppa in Grecia/Egeo BA	aDNA Egeo III-II mill.	100+ individui	5-8 anni
Bassa	Filogenesi bayesiana IE (dati freschi)	Corpus linguistico+BEAST2 aggiornato	Corpus esistente	2-4 anni

APPENDICI

Appendice A — Tavola sinottica dei profili genomici delle culture steppiche

Cultura	Periodo	Area	N indiv. aDNA	EHG%	CHG%	EEF%	WHG%	Aplogruppo Y dom.
Khvalynsk	4700-3800 a.C.	Medio Volga	12	82	15	2	1	R1b-V1636, Q
Sredny Stog	4500-3500 a.C.	Dnepr-Don	8	68	28	3	1	R1a-M417, I2
Maykop	3700-3000 a.C.	Caucaso Nord	22	22	62	12	4	J2, G2, R1b
Yamnaya Samara	3300-2700 a.C.	Volga inf.	65	52	43	4	1	R1b-M269, I2a
Yamnaya Dnepr	3300-2700 a.C.	Ucraina S	38	46	48	5	1	R1b-M269
Yamnaya Caucaso	3100-2600 a.C.	Caucaso N	18	33	60	6	1	R1b, J
Afanasievo	3300-2500 a.C.	Altai	24	62	34	3	1	R1b-M269
Corded Ware	2900-2300 a.C.	Europa C	148	40	28	22	10	R1a-M417
Bell Beaker Iberia	2500-1800 a.C.	Penisola Ib.	94	22	17	54	7	R1b-P312
Bell Beaker UK	2400-1800 a.C.	Isole Brit.	78	44	28	21	7	R1b-P312
Sintashta	2100-1800 a.C.	Urali S	48	57	33	9	1	R1a-Z93
Andronovo	1800-1400 a.C.	Kazakhstan	35	60	31	8	1	R1a-Z93

Appendice B — Derivazione della statistica ABBA-BABA (D-statistic): dimostrazione formale

La derivazione formale della D-statistic parte dal modello coalescente sotto un albero a quattro popolazioni. Sia T l'albero con topologia ((P1,P2),P3,O) e siano t_12 < t_123 i tempi di coalescenza (in unita' di 2N generazioni). Per un SNP biallelico con allele ancestrale "A" e derivato "B", i pattern di sito possono essere classificati in termini delle sequenze (P1,P2,P3,O):

ABBA: (A,B,B,A) — derivato condiviso da P2 e P3

BABA: (B,A,B,A) — derivato condiviso da P1 e P3

BBAA: (B,B,A,A) — derivato condiviso da P1 e P2 (non informativo)

Sotto l'albero senza flusso genico, le probabilita' di ciascun pattern dipendono dai tempi di ramificazione e dalla taglia efficace. Per la topologia ((P1,P2),P3,O), il tempo interno nel quale una mutazione puo' segregare asimmetricamente tra P1/P2 e P3 dipende da t_123 - t_12. Se la mutazione sorge in questo intervallo su un lignaggio che poi va a P3 (producendo ABBA) o su un lignaggio che poi va a P1 (producendo BABA), i due eventi sono equiprobabili per simmetria (entrambe le linee ancestrali di P1 e P2 sono fungibili nell'intervallo t_12 < t < t_123).

La differenza ABBA - BABA ha valore atteso zero per simmetria, e il suo valore osservato non zero fornisce il test di flusso genico:

$$D = \frac{\sum_j \left[f\left(ABBA_j\right) - f\left(BABA_j\right) \right]}{\sum_j \left[f\left(ABBA_j\right) + f\left(BABA_j\right) \right]}$$

dove f(ABBA_j) = p_{Aj}*(1-p_{Bj})*p_{Cj}*(1-p_{Oj}) per il peso del sito j ai quattro campioni P1,P2,P3,O rispettivamente. La standardizzazione tramite block jackknife sui blocchi da 5 cM (distanza genetica) tiene conto della correlazione per linkage disequilibrium.

Per il caso di popolazioni individuali piuttosto che medie di popolazione (come negli studi di aDNA con bassi numeri di campioni), il metodo "pseudo-haploid" campiona un singolo allele per individuo a ogni SNP, riducendo il bias da eterozigosita'. La varianza della statistica D viene stimata come:

$$Var(D) = \frac{Var(\sum ABBA - BABA)}{\left(\sum ABBA + BABA\right)^2}$$

$$\sigma_D^2 \approx \frac{1}{B}\sum_{b=1}^{B}\left(D_b - \bar{D}\right)^2 \quad [block\ jackknife]$$

dove D_b e' il valore D calcolato sul blocco b (delete-one jackknife), e B e' il numero di blocchi genomici. Il numero di blocchi tipicamente usato e' 1233 (per il genoma umano diviso in blocchi da ~5 cM). Il Z-score = D/sqrt(Var(D)) segue asintoticamente una distribuzione normale standard, permettendo test di significativita' bilaterali.

Appendice C — Filogenesi linguistica bayesiana: il modello BEAST per le lingue IE

La filogenesi bayesiana delle lingue indoeuropee applica metodi sviluppati originalmente per la filogenetica molecolare al problema della ricostruzione dell'albero genealogico delle lingue IE e della datazione dei suoi nodi. Il software BEAST (Bayesian Evolutionary Analysis Sampling Trees) di Drummond e Rambaut utilizza metodi MCMC (Markov Chain Monte Carlo) per campionare dallo spazio degli alberi filogenetici pesato dalla verosimiglianza dei dati.

Il modello di evoluzione lessicale utilizzato piu' comunemente e' il modello Stochastic Dollo (Nicholls e Gray 2006): ogni carattere (cognate set, ovvero gruppo di parole che derivano da una stessa radice PIE) viene acquisito una sola volta nella storia dell'albero (innovazione unica) e poi puo' essere perso in modo stocastico. Formalmente, sia q_01 il tasso di acquisizione e q_10 il tasso di perdita di un carattere, con q_01 molto piccolo (acquisizione rara) e q_10 variabile per carattere. La verosimiglianza di osservare il pattern di presenza/assenza del carattere nelle lingue fogliari dato l'albero T e i parametri di tasso si calcola con l'algoritmo di Felsenstein (pruning algorithm):

$$L(T,\theta) = P(D|T,\theta) = \prod_{c}\sum_{s_{root}} \pi\left(s_{root}\right)\prod_{edges} P\left(s_{child}|s_{parent}, t_e, \theta\right)$$

dove la somma e' sul stato del carattere alla radice, pi(s) e' la distribuzione stazionaria, e P(s_child | s_parent, t, q) e' la probabilita' di transizione calcolata esponenzialmente dalla matrice di tasso Q = [[-q_01, q_01],[q_10, -q_10]]:

$$P(t) = \exp(Q*t)$$

Il modello di orologio molecolare usato per IE e' tipicamente un orologio rilassato (relaxed clock) che permette ai tassi di evoluzione di variare tra i rami secondo una

distribuzione log-normale. I vincoli di calibrazione ("constraints") sono forniti da attestazioni storiche delle lingue: il greco miceneo (~1350 a.C.), il latino arcaico (~200 a.C.), il vedico (~1200 a.C.), l'ittita (~1650 a.C.) forniscono lower bounds sui tempi di separazione dei rispettivi rami.

I risultati di Chang et al. (2015) con 207 cognate sets e 103 lingue IE forniscono una data per la radice dell'albero IE di 6700-4000 a.C. (95% HPD) con il metodo MCMC — coerente con l'ipotesi kurganica ma con intervallo di confidenza ampio che non esclude completamente l'ipotesi anatolica. Heggarty et al. (2023) con un dataset piu' grande (161 lingue, 4500 cognate items) ottengono 8120-5940 a.C. (95% HPD), risultato che ha riacceso il dibattito perche' la stima e' piu' alta del previsto e parzialmente compatibile con la neolitizzazione. Tuttavia, gli autori stessi notano che la posizione basale dell'anatolico e la struttura dell'albero sono piu' consistenti con l'ipotesi kurganica che con quella anatolica.

Appendice D — Il modello qpGraph: ricostruzione dell'albero con flusso genico

Il software qpGraph, sviluppato da Patterson et al. (2012) e integrato in ADMIXTOOLS, permette di testare modelli espliciti dell'albero delle popolazioni con flusso genico, stimando contemporaneamente la topologia dell'albero, le lunghezze dei rami (in unita' di deriva genetica) e i pesi dei rami di migrazione. Il modello e' specificato dall'utente come un grafo diretto aciclico (DAG) con nodi interni che rappresentano popolazioni ancestrali ipotetiche, foglie che rappresentano le popolazioni osservate, e archi di migrazione che permettono flusso genico tra rami.

La funzione obiettivo e' il "worst f4 residual" — la massima deviazione tra le statistiche f4 osservate e quelle predette dal modello, normalizzata per l'errore standard. Un modello si considera adeguato se il worst f4 residual e' < 3 (in unita' di deviazioni standard), e si considera ottimale se e' < 2. Per confrontare modelli alternativi (ad es. rotta balcanica versus rotta caucasica per l'anatolico), si confrontano i residui dei due modelli e si preferisce quello con residui minori.

La difficolta' computazionale del metodo qpGraph e' che lo spazio dei modelli e' enorme: anche per un numero modesto di popolazioni (10-15), il numero di possibili

topologie con 0-3 rami di migrazione e' astronomico. L'esplorazione esaustiva non e' fattibile; si utilizzano invece strategie di ottimizzazione locale (gradient descent nel numero dei parametri di ramo) o algoritmi evolutivi che esplorano lo spazio dei modelli in modo stocastico. Il software AdmixtureBayes (Nielsenetal 2023) implementa un approccio RJMCMC (Reversible Jump MCMC) che campiona dallo spazio dei modelli in modo proporzionale alla loro verosimiglianza, producendo distribuzioni posteriori sulle topologie e sui pesi di migrazione che quantificano formalmente l'incertezza strutturale del modello.

LE ORIGINI DEI PALEOINDOEUROPEI

PARTE SECONDA: LINGUISTICA STORICA, MORFOLOGIA E DISPERSIONE DELLA FAMIGLIA INDOEUROPEA

EDIZIONE CRITICA — FONOLOGIA, FILOGENESI LINGUISTICA E MODELLI DI AMMIXTURE

Silvano Salvador

CAPITOLO III — LA RICOSTRUZIONE FONOLOGICA DEL PROTO-INDOEUROPEO: INVENTARIO, LEGGI E STRUTTURA

III.1 Il sistema consonantico del PIE: le occlusive e il problema delle laringali

III.1.1 L'inventario delle occlusive: la disputa tra il sistema a due e a tre serie

La ricostruzione del sistema consonantico del Proto-Indoeuropeo e' uno dei risultati piu' maturi della linguistica storica ottocentesca e novecentesca, e al tempo stesso uno degli argomenti piu' controversi della linguistica teorica contemporanea. Il sistema tradizionalmente ricostruito, articolato a partire dalle corrispondenze sistematiche osservabili nelle lingue IE storicamente attestate, e' un sistema a tre serie di occlusive: occlusive sorde (*p, *t, *k, *kw), occlusive sonore (*b, *d, *g, *gw) e occlusive sonore aspirate (*bh, *dh, *gh, *gwh). Questa tripartizione, elaborata sistematicamente da Karl Brugmann nel Grundriss der vergleichenden Grammatik del 1886, si fonda sulla regolarita' delle corrispondenze fonetiche tra sanscrito, greco, latino, germanico e altre lingue IE: le stesse posizioni consonantiche nelle stesse radici corrispondono sistematicamente tra le varie lingue in accordo con equazioni di corrispondenza precise.

La corrispondenza fondamentale che documenta le tre serie e' la seguente. La radice PIE *bher- (portare), che appare come bhara- in sanscrito, phero in greco, fero in latino, bara in gotico antico, mostra una consonante iniziale che corrisponde a bh in sanscrito, ph in greco (con successiva spirantizzazione a f), f in latino, b in germanico. Questa corrispondenza e' l'esito atteso se la consonante originaria era *bh (occlusiva sonora aspirata): il sanscrito conserva l'aspirazione; il greco spira l'occlusiva (bh > ph); il latino spirantizza ph > f; il germanico perde l'aspirazione ma conserva la sonorita'. Le tre serie si comportano in modo regolare e diversificato in ciascuna lingua figlia, permettendo la ricostruzione della serie originaria con alto grado di certezza.

Il problema teorico principale del sistema tradizionale e' l'asimmetria della distribuzione dei fonemi. L'occlusiva labiale sorda *p e' estremamente comune nelle radici PIE ricostruite; l'occlusiva labiale sonora *b e' eccezionalmente rara — di fatto,

non esiste un singolo morphema PIE ricostruito che contenga esclusivamente *b senza che la sua esistenza sia controversa. Questa rarezza di *b contrasta sia con la frequenza tipologicamente normale delle occlusive sonore nelle lingue del mondo sia con la frequenza delle occlusiva dentale *d e velare *g nel PIE ricostruito. La rarezza di *b ha spinto alcuni linguisti (la teoria delle glottalizzate, proposta da T.V. Gamkrelidze e V.V. Ivanov negli anni Settanta e sviluppata da Paul Hopper) a proporre una ricostruzione radicalmente diversa del sistema occlusivo del PIE.

La teoria delle glottalizzate (nota anche come "glottalic theory") propone di sostituire il sistema tradizionale {sorde, sonore, sonore aspirate} con il sistema {sorde, glottalizzate/eiezettive, aspirate sorde}. In questa riscrittura, la serie tradizionalmente ricostruita come *b, *d, *g diventa una serie di consonanti glottalizzate (occlusive eiezettive o iniettive) *p', *t', *k'; le sorde tradizionali *p, *t, *k diventano aspirate sorde *ph, *th, *kh; e le sonore aspirate tradizionali *bh, *dh, *gh diventano sonore *b, *d, *g. Il vantaggio fonologico della teoria delle glottalizzate e' di produrre un sistema consonantico tipologicamente piu' plausibile (i sistemi con glottalizzate ma senza vere sonore sono documentati nelle lingue del Caucaso, dell'America e dell'Africa) e di spiegare naturalmente la rarezza di *b (le glottalizzate labiali sono universalmente rare nelle lingue del mondo). Il vantaggio geografico e' la coerenza con la proposta di un Urheimat caucasico di Gamkrelidze e Ivanov stessi. Tuttavia, la teoria delle glottalizzate non ha raggiunto consenso nella comunita' linguistica, principalmente per la difficolta' di derivare gli esiti attestati nelle lingue IE dal sistema glottalizzato in modo altrettanto naturale quanto il sistema tradizionale.

Termine PIE	**Sanscrito**	**Greco**	**Latino**	**Gotico**	**Lituano**	**Ittita**	**Significato**
*pater-	pita	pater	pater	fadar	teva	attas	padre
*mater-	matar	meter	mater	aidem	mote	anna	madre
*bhrater-	bhratar	phrater	frater	brothar	brolis	pankur	fratello
*dwo-	dvi	duo	duo	twai	du	ta	due
*treyes	tri	treis	tres	threis	trys	tri	tre
*kwekwlos	cakra	kyklos	circus	--	kyklas	--	cerchio/ruota
*genu-	janu	gony	genu	kniu	kelis	genu	ginocchio

Termine PIE	Sanscrito	Greco	Latino	Gotico	Lituano	Ittita	Significato
*ekwos	asva	hippos	equus	aihva	asva	assuwa	cavallo
*wlkwos	vrka	lykos	lupus	wulfs	vilkas	ulippas	lupo
*ghosti-	--	xenos	hostis	gast	--	--	straniero/ospite

III.1.2 La teoria laringale: de Saussure, Kurilowicz e la conferma hittitica

La teoria laringale costituisce il piu' spettacolare caso di previsione confermata nella storia della linguistica storica — e al tempo stesso uno dei contributi intellettuali piu' originali dell'Ottocento scientifico. Nel 1879, Ferdinand de Saussure, all'epoca ventunenne studente a Lipsia, pubblica il Memoire sur le systeme primitif des voyelles dans les langues indo-europeennes. In questo testo — composto in un momento in cui il sanscrito, il greco, il latino e il gotico erano le lingue IE meglio conosciute e nessuna lingua anatolica era ancora decifrata — de Saussure propone che il sistema vocalico del PIE debba essere integrato dall'ipotesi di elementi consonantici (da lui chiamati "coefficienti sonantici") la cui esistenza e' richiesta dalle corrispondenze sistematiche nelle alternanze vocaliche dei morfemi. De Saussure identifica due coefficienti principali (oggi denominati convenzionalmente H1, H2, H3) e ne specifica le proprieta' mediante le loro interazioni con le vocali adiacenti: H2 colorisce la vocale adiacente in /a/, H3 in /o/, H1 lascia la vocale inalterata.

La conferma spettacolare viene nel 1927, quando Jerzy Kurilowicz dimostra che il fricativo laringale h dell'ittita — un segno cuneiforme che i filologi avevano notato senza capirne la funzione sistematica — corrisponde esattamente alle posizioni in cui de Saussure aveva postulato la presenza dei suoi coefficienti cinquant'anni prima. L'ittita, una lingua IE documentata ma sconosciuta ai tempi di de Saussure, ha conservato fossili delle laringali in posizioni dove tutte le altre lingue IE le hanno perse, esattamente come previsto dall'analisi comparativa. Questa coincidenza — la previsione di un elemento linguistico basata su ragionamento strutturale puramente interno, confermata quarantotto anni dopo da evidenza esterna — e' paragonabile alla previsione del pianeta Nettuno da parte di Le Verrier o della previsione della forma del DNA basata sui dati di diffrazione di Franklin.

Le laringali del PIE si ricostruiscono convenzionalmente come tre fonemi: *H1 (probabilmente fricativa glottidale /h/), *H2 (probabilmente fricativa uvulare sorda /x/ o faringale /X/), e *H3 (probabilmente fricativa laringale sonora o fricativa labiovelare /xw/). Le loro interazioni con le vocali adiacenti producono gli "effetti di colorazione" vocalica che motivano la teoria. Prima di vocale, le laringali sono trasparenti ma modificano la vocale successiva: *H1e > e, *H2e > a, *H3e > o. Dopo consonante e prima di consonante (posizione intersillabica), le laringali divengono vocali "sonanti" con valori e, a, o rispettivamente — spiegando la distribuzione altrimenti irregolare delle alternanze e/a/o nelle radici PIE. In posizione iniziale prima di vocale, la laringale e' conservata in ittita come h- (attestazione diretta); in tutte le altre lingue IE e' perduta con compensazione vocalica.

Regola laringale	PIE ricostruito	Ittita	Greco	Sanscrito	Latino	Esito
H1 + e > e	*H1esti	eszi	esti	asti	est	'e' (essere)
H2 + e > a	*H2nemo-	aniur	anemos	anila	animus	vento/spirito
H3 + e > o	*H3okwi-	aku-	opsis	aksi	oculus	occhio
VH vocale > vocale lunga	*dH2-to-	--	dotos	ditas	datus	dato
CRHC > CaRC (H2)	*pH2ter-	attas	pater	pitar	pater	padre
H in posizione iniziale	*H2erti-	harti	arti	rti	ars	arte/abilita
H3 > labiov. contesto	*H3owi-	hawi	ovis	avi	ovis	pecora

III.1.3 Le leggi fonetiche come strumento epistemico: Grimm, Verner e Brugmann

Le leggi fonetiche — regolarita' sistematiche nelle corrispondenze tra i suoni di lingue imparentate — costituiscono lo strumento epistemico fondamentale della linguistica storica. La loro formulazione come leggi senza eccezioni (Ausnahmslosigkeit der Lautgesetze, il principio dei neogrammatici) rappresenta la svolta metodologica che ha trasformato la glottologia da disciplina speculativa in scienza rigorosa. La prima grande legge fonetica sistematicamente formulata e' la

Legge di Grimm (Jakob Grimm, Deutsche Grammatik, 1819-1837), che descrive il mutamento consonantico sistematico che separa il germanico da tutte le altre lingue IE.

La Legge di Grimm opera in tre fasi consecutive, producendo uno "scivolamento" ciclico delle consonanti. Nella prima fase, le occlusive sonore aspirate del PIE (*bh, *dh, *gh) diventano fricative sonore (b, d, g > v/b, d/th, g/gh in gotico antico). Nella seconda fase, le occlusive sonore non aspirate del PIE (*b, *d, *g) diventano sorde (b, d, g > p, t, k in germanico). Nella terza fase, le occlusive sorde del PIE (*p, *t, *k) diventano fricative sorde (p, t, k > f, th, x/h in germanico). Il risultato complessivo e' un cambiamento sistematico di tre serie in un ciclo, descrivibile formalmente come:

*bh, *dh, *gh -> beta, delta, gamma (poi f, th, h in posiz. iniziale)

*b, *d, *g -> p, t, k

*p, *t, *k -> f, th, h (fricative sorde)

Le eccezioni apparenti alla Legge di Grimm — che sembravano minacciare il principio di regolarita' — trovano spiegazione nella Legge di Verner (Karl Verner, 1875). Verner osservo' che le fricative sorde germaniche derivate da *p, *t, *k si sonorizzano in b, d, g quando nel PIE il tono non cadeva sulla sillaba immediatamente precedente: il contesto di sonorizzazione e' determinato dall'accento del PIE, non dalla struttura segmentale germanica. Questo significa che le cosiddette "eccezioni" a Grimm erano in realta' regolari — regolate da un condizionamento accentuale che il germanico aveva nel frattempo eliminato (il germanico ha fissato l'accento sulla sillaba iniziale della radice, rendendo invisibile il condizionamento originario). La Legge di Verner e' quindi non un'eccezione ma una seconda legge che interagisce con Grimm, confermando ulteriormente la regolarita' dei mutamenti fonetici.

La Legge di Brugmann (o "legge di allungamento") descrive un processo specifico del greco antico: la vocale breve in sillaba aperta davanti a sonorale (nasale + fricativa sorda) si allunga. Sebbene di portata piu' limitata rispetto a Grimm e Verner, la Legge di Brugmann e' esemplificativa della precisione con cui i neogrammatici specificavano i condizionamenti ambientali dei mutamenti fonetici — un livello di dettaglio che anticipa la moderna fonologia generativa. La codifica formalizzata di queste leggi nel XIX secolo ha fornito al comparativismo

indoeuropeo un sistema di regole predittive paragonabile per rigore, se non per formalizzazione matematica, alle leggi fisiche.

Legge fonetica	Autore/Anno	Lingua	Descrizione sintetica	Esempio (PIE > esito)
Legge di Grimm	J. Grimm, 1822	Germanico	Scivolamento consonantico tripartito	*pater > fadar (gotico)
Legge di Verner	K. Verner, 1875	Germanico	Sonorizzazione condizionata dall'accento PIE	*bhrater > brodher
Ruki Rule	vari, XIX sec.	Slavo/Baltico/Indo-ir.	s > sh dopo r,u,k,i	*weso > vedico vah
Legge di Brugmann	K. Brugmann, 1876	Greco	Allungamento vocalico in sill. aperta	*gen- > genos
Legge di Bartholomae	C. Bartholomae, 1882	Indo-iranico	Assimilazione regressiva aspirate	*gadh+ta > indiana gadha
Legge di Sievers	E. Sievers, 1878	Indoeur. generale	Alternanza i/y dipendente dalla sillaba	*dyeu/dieus
Legge di Stang	C. Stang, 1966	Balto-slavo	Metatonia nelle lengature tonali	acc. baltico mobile
Legge di Dybo	V. Dybo, 1961	Slavo	Spostamento accentuale in pre-baltico	alternanze acc. slavo

III.2 Il sistema vocalico e l'apofonia: fondamenti e applicazioni ricostruttive

III.2.1 Il sistema apofonico del PIE: gradi e ablaut

L'apofonia — la variazione sistematica della qualita' e della quantita' vocalica nelle radici e nei suffissi del PIE — e' uno dei fenomeni morfologici piu' caratteristici e pervasivi della famiglia IE. Nelle lingue IE arcaiche, la stessa radice lessicale appare con vocali diverse in forme grammaticali diverse: la radice che significa "portare" appare come *bher- (grado e), *bhor- (grado o), e *bhr- (grado zero, con sillabificazione della sonorale r). Questa variazione non e' irregolare ma segue schemi precisi, detti "gradazioni apofoniche" o "ablaut", che sono ricostruibili in forma sistematica per il PIE.

Il sistema apofonico del PIE distingue tradizionalmente tre gradi fondamentali. Il grado e (o grado normale): la radice contiene la vocale *e; es. *bher-, *leg-, *ped-. Il grado o (o grado ridotto timbrico): la vocale *e e' rimpiazzata da *o; es. *bhor-, *log-, *pod-. Il grado zero (o grado ridotto): la vocale e' eliminata e la sonorale adiacente (m, n, r, l, i, u) assume funzione sillabica; es. *bhr-, *lg-, *pd-. A questi tre gradi si aggiungono gradi allungati (grado allungato e = *e: e grado allungato o = *o:), usati in certi paradigmi morfologici, in particolare nei vocativi singolari e in alcune forme verbali. L'alternanza sistematica tra i gradi e' morfologicamente condizionata: tipicamente il grado e appare nelle forme con accento sulla radice, il grado zero nelle forme con accento sul suffisso o sulla desinenza, e il grado o in certe funzioni grammaticali specifiche (nomina agentis, perfetto verbale).

La formalizzazione matematica dell'apofonia e' stata proposta da varios linguisti nel quadro della fonologia generativa (Kiparsky 1973, Marantz 1982). Sia R una radice e S il suo slot vocalico; il meccanismo apofonico puo' essere rappresentato come una regola morfofonologica che mappa il grado apofonico (un tratto morfologico) sul timbro vocalico:

Grade(e): /CeC/ -> CeC [grado normale]

Grade(o): /CeC/ -> CoC [grado o, metafonia]

Grade(0): /CeC/ -> CC [elisione con sillabificazione sonorale]

Grade(e:): /CeC/ -> Ce:C [grado allungato]

Questa formalizzazione cattura l'intuizione che l'apofonia e' un processo morfofonologico governato da regole, non una variazione irregolare. La parallelo con processi morfologici analoghi in altre lingue (Ablaut nell'inglese sing/sang/sung, i/a/u, e' un residuo fossilizzato dello stesso meccanismo PIE) illustra la persistenza di strutture apofoniche nelle lingue IE moderne anche dopo la perdita del sistema morfologico originario.

Radice PIE	Grado e (normale)	Grado o	Grado zero	Grado allungato	Significato
*bher-	*bher-o (porto)	*bhor- (portatore)	*bhr-to- (portato)	*bhe:r- (raro)	portare
*leg-	*leg- (raccogliere)	*log- (raccolta)	*lg-to- (raccolto)	*le:g-	raccogliere/leggere

Radice PIE	Grado e (normale)	Grado o	Grado zero	Grado allungato	Significato
*ped-	*ped- (piede)	*pod- (podio)	*pd-os (genitivo)	*pe:d-	piede/cadere
*gwen-	*gwen- (donne)	*gwon-	*gwn-	*gwe:n-	donna
*deik-	*deik- (mostrare)	*doik-	*dik-to-	*de:ik-	mostrare/dire
*gen-	*gen- (nascere)	*gon- (nato)	*gn-to- (nato)	*ge:n-	nascere/genere
*weid-	*weid- (vedere)	*woid- (sa)	*wid- (visto)	*we:id-	vedere/sapere

III.2.2 La morfologia nominale del PIE: il sistema degli otto casi

Il sistema dei casi del PIE rappresenta uno degli edifici morfologici piu' elaborati ricostruibili per una lingua preistorica. La ricostruzione standard distingue otto casi nel singolare (con variazioni nel duale e nel plurale): nominativo, vocativo, accusativo, genitivo, dativo, ablativo, strumentale e locativo. Non tutte le lingue IE storicamente attestate conservano la distinzione completa degli otto casi: il latino ne conserva sei (fondendo ablativo e strumentale nel cosiddetto ablativo, e perdendo il locativo come caso produttivo); il greco ne conserva cinque; le lingue balto-slave sono quelle che meglio conservano la struttura originaria con sette-otto casi. L'ittita mostra un sistema di soli cinque-sei casi nella sua fase documentata, ma con tracce di forme piu' arcaiche.

La desinenza di ciascun caso e' ricostruita a partire dalle corrispondenze regolari nelle lingue IE. Per i temi in -o- (la classe nominale piu' numerosa, comprensiva dei nomi maschili e neutri), le desinenze ricostruite del PIE sono le seguenti:

Caso	Desinenza PIE (sg.)	Sanscrito	Greco	Latino	Lituano	Gotico	Funzione
Nominativo	*-os	-as	-os	-us	-as	-s	Soggetto
Vocativo	*-e	-a	-e	-e	-e	-	Invocazione
Accusativo	*-om	-am	-on	-um	-a	-	Oggetto diretto
Genitivo	*-osyo	-asya	-oio/-ou	-i	-o	-is	Possesso/partitivo
Dativo	*-oi	-aya	-oi	-o	-ui	-a	Destinatario

Caso	Desinenza PIE (sg.)	Sanscrito	Greco	Latino	Lituano	Gotico	Funzione
Ablativo	*-o:d	-at	(=gen.)	-od	(=gen.)	-	Allontanamento
Strumentale	*-oh1	-a	(=dat.)	(=abl.)	-u	(=dat.)	Mezzo/accomp.
Locativo	*-oi / *-ei	-e	(=dat.)	-i	-ame	-	Luogo

La ricostruzione delle desinenze plurali e duali presenta complicazioni aggiuntive, in quanto non tutte le lingue IE conservano il duale e alcune hanno semplificato i plurali in modo divergente. Le desinenze del nominativo plurale *-oes/*-o:s (con apofonia della desinenza) e del genitivo plurale *-om (attestato in sanscrito -am, greco -on, latino -um, celtico -on) sono tra le corrispondenze piu' robuste e forniscono conferma della profondita' genealogica della famiglia IE. La genesi storica del sistema degli otto casi e' oggetto di dibattito: una linea di ricerca sostiene che il sistema dei casi IE si sia sviluppato da un sistema piu' antico con distinzione ternaria soggetto/oggetto/obliquo, analoga a quella riscontrabile in lingue come l'austronesiano; un'altra linea vede il sistema degli otto casi come eredita' di un sistema originario ancora piu' ricco.

III.3 Il lessico culturale ricostruito e le implicazioni storico-geografiche

III.3.1 Il lessico del bestiame, del carro e del guerriero: datazione indipendente della dispersione

Una delle metodologie piu' fruttuose — e al tempo stesso piu' dibattute — nella ricerca sull'Urheimat e' la Worterscheinungsgeschichte, ovvero lo studio delle condizioni di esistenza del lessico PIE ricostruito come strumento per inferire il contesto geografico, culturale e cronologico dei parlanti. L'idea di base e' semplice: se il PIE ricostruisce un termine con significato X, i parlanti del PIE conoscevano X; la distribuzione geografica e la datazione di X nell'evidenza archeologica fornisce quindi un vincolo indipendente sull'Urheimat. Il procedimento e' intuitivamente valido ma metodologicamente ricco di insidie: (a) un termine ricostruito potrebbe avere significato diverso da quello inferito; (b) un termine potrebbe essere stato

innovato indipendentemente in rami diversi (prestiti paralleli); (c) l'assenza di un termine ricostruito in un ramo non significa necessariamente che i parlanti non conoscessero il referente (la perdita lessicale e' comune).

Con queste cautele, il lessico culturale PIE offre vincoli cronologici e geografici notevoli. Il termine piu' importante per la cronologia assoluta e' il lessico del carro: il PIE ricostruisce *kwekwlos (ruota, da *kwel- girare; cfr. sanscrito cakra, greco kyklos, inglese wheel), *H2eks-le- (asse, cfr. latino axis, lituano asis), *H3ibh-r (mozzo, cfr. ittita hiparas?), e potenzialmente *wergh- (per il carrello/carro). La presenza di questa terminologia nel PIE implica che i parlanti del PIE conoscevano la ruota e probabilmente il carro con ruota raggiata. La datazione archeologica delle prime ruote raggiando e' di circa 2500-2000 a.C. nelle steppe; le ruote piene (disco) compaiono prima, intorno a 3500 a.C. in Mesopotamia e 3000 a.C. nelle steppe. Se il termine ricostruito indica specificamente la ruota raggiata (il che e' incerto), la dispersione del PIE non puo' essere anteriore al 2500 a.C. — un vincolo che penalizza l'ipotesi anatolica nella sua forma classica (che pone la dispersione PIE nel Neolitico, 7000-5000 a.C.) e che e' compatibile con l'ipotesi kurganica.

Il lessico della guerra e del guerriero nel PIE ricostruisce un sistema semantico elaborato: *H2ner- (guerriero/uomo vitale, cfr. greco aner, irlandese nar, osseto nar), *H3ek-s- (asse/arma, cfr. greco axine), *H2ewsi- (oro, cfr. sanscrito ayas, latino aurum — contrastato con *H3ayes- per bronzo/rame, cfr. latino aes, vedico ayas). La presenza di termini per metalli lavorate e in particolare del bronzo nel lessico PIE e' consistent con la datazione kurganica: l'eta' del Bronzo nelle steppe inizia intorno al 3500-3000 a.C. con i tellurosi di Maykop, mentre la diffusione del bronzo nel senso moderno (lega di rame e stagno) si situa intorno al 3000-2500 a.C. nell'Europa orientale. Il lessico PIE per i metalli e' tuttavia problematico: molti termini mostrano distribuzione irregolare tra i rami e potrebbero essere prestiti o convergenze tardive.

Campo semantico PIE	Radice ricostruita	Attestazioni	Implicazione cronologica	Implicazione geografica
Bestiame bovino	*gwou-	sanscr. gaus, greco bous, latino bos	Conoscenza bue domestico	Compatibile con steppe e Anatolia
Cavallo	*H1ekwo-	sanscr. asva, greco hippos, latino equus	Conoscenza cavallo domestico	Steppe: domesticazione ~3500 a.C.

Campo semantico PIE	Radice ricostruita	Attestazioni	Implicazione cronologica	Implicazione geografica
Pecora/lana	*H3owi-	sanscr. avi, greco ovis, latino ovis	Conoscenza ovino domestico	Largo: Vicino Oriente + steppe
Ruota	*kwekwlo-	sanscr. cakra, greco kyklos, gall. cucol	Carro a ruote noto	Steppe/Caucaso: ~3000-2500 a.C.
Asse del carro	*H2ekso-	latino axis, lituano asis, greco axon	Sistema carriggio noto	Steppe: ~3000 a.C.
Miele/idromele	*melit-	greco meli, latino mel, gotico milith	Apicoltura nota	Nord: api presenti in Europa e steppe
Neve/inverno	*sneigwh-	sanscr. sneha, greco nipha, gotico snaiws	Clima continentale con neve	Steppe o nord Europa
Salmone	*laks-	gotico lahs, lituano lasis, russo losos'	Conoscenza salmone	Fiumi del nord (non Anatolia)
Betulla	*bheH1go-	russo bjoza, sanscr. bhurja, lituano berzas	Albero boreale noto	Steppe settentrionali/Europa nord
Signore/re	*H3reg-	latino rex, sanscr. raja, celtico rix	Struttura regale nota	Compatibile con eta' del Bronzo

III.3.2 Il lessico della parentela e le strutture sociali dei Proto-IE

Il lessico della parentela del PIE e' uno dei campi semantici piu' ricchi e meglio ricostruibili, e fornisce indicazioni dirette sulla struttura sociale e sulla organizzazione familiare dei parlanti del PIE. La terminologia di parentela IE segue il sistema classificatorio noto come "sistema biforcato-colaterale" o, nella classificazione di Morgan, come sistema "descrittivo": i parenti sono classificati con termini distinti a seconda del sesso del genitore intermedio, producendo una distinzione sistematica tra parenti paterni e materni che e' assente nei sistemi "lineare" delle lingue romanze moderne (dove il termine "zio" comprende sia lo zio paterno che lo zio materno).

In particolare, il PIE ricostruisce una serie di termini specifici per i parenti del lato paterno: *H2awos (nonno paterno, cfr. latino avus, gotico awi), *H2ewH2os (nonna paterna, cfr. latino avia), *pHter kHH1nis (fratello del padre = zio paterno, cfr.

vedico pitruvya, latino patruus, greco patroios), e probabilmente *swesor (sorella, con derivati distinti per sorella del padre e sorella della madre nelle singole lingue figlie). Questo sistema di differenziazione sistematica del lato paterno e' coerente con una struttura sociale patrilineare — nella quale il lignaggio si trasmette per linea maschile — che e' confermata dai dati archeologici e genomici: le culture delle steppe mostrano aplogruppi Y fortemente omotetici (pochi aplogruppi paterni dominanti, come R1b e R1a) con alta diversita' mitocondriale, pattern genomico che indica una struttura patrilineare con mobilita' delle donne tra i gruppi.

Il lessico PIE ricostruisce anche i termini fondamentali della gerarchia sociale: *H3reg- (re, da *H3reg- guidare/dirigere in linea retta, cfr. latino rex, sanscrito raja, celtico *rix), *H2wid-tora- (sapiente/veggente, cfr. gotico witan), e probabilmente *dH3-to- (dato/legge) da *deH3- (dare). La presenza di un termine per "re" nel PIE ricostruito e' compatibile con una struttura sociale gerarchica, ma non indica necessariamente un sistema politico centralizzato dell'eta' del Bronzo: il *H3reg- del PIE era verosimilmente un capo tribale-pastorale, non un monarca di tipo vicino-orientale. Questa inferenza e' coerente con le evidenze archeologiche delle steppe, che mostrano strutture funerarie kurganiche con differenziazione del corredo (indicando gerarchia di status) ma senza le mega-strutture monumentali delle civilta' idrauliche.

Termine PIE	Ricostruzione	Sanscrito	Latino	Greco	Gotico	Lituano	Significato
Padre	*pHter	pitar	pater	pater	fadar	teva	padre
Madre	*meHter	matar	mater	meter	aidem	mote	madre
Fratello	*bhrH2ter	bhratar	frater	phrater	brothar	brolis	fratello
Sorella	*swesor	svasar	soror	eor	swistar	sesuo	sorella
Figlio	*suHnu-	sunu	--	--	sunus	sunus	figlio
Figlia	*dhuH1ter-	duhitar	--	thygater	dauhtar	dukte	figlia
Nonno pat.	*H2awos	--	avus	--	awii	--	nonno (paterno)
Zio paterno	*pHter-kwHnis	pitruvya	patruus	patroios	--	--	fratello del padre

Termine PIE	Ricostruzione	Sanscrito	Latino	Greco	Gotico	Lituano	Significato
Suocero	*sweH1kuro-	svashura	socer	hekuros	swaihra	--	padre del coniuge
Cognato	*daH2iwer-	devar	--	daer	zeihor	--	fratello del marito

CAPITOLO IV — LA DISPERSIONE INDOEUROPEA: CULTURE, GENOMICA E CRONOLOGIA COMPARATA

IV.1 La cultura Sintashta e il problema dell'Indo-iranico

IV.1.1 Sintashta: la culla del proto-Indo-iranico e la rivoluzione militare delle steppe

La cultura Sintashta — identificata nei siti degli Urali meridionali (Russia sud-orientale, Kazakhstan settentrionale) e datata approssimativamente tra 2100 e 1800 a.C. — rappresenta uno dei fenomeni piu' significativi e meglio documentati della preistoria eurasiatica. Il complesso di Sintashta comprende circa venti "paesi delle steppe" (in russo степные страны, Stepnye Strany), insediamenti urbani di piccole dimensioni (tipicamente 2-8 ettari) con pianta ovale o circolare, mura di protezione in terra o in legno, edifici regolari a sviluppo rettangolare, e cimiteri con kurgan monumentali che contengono le prime sepolture di carri con ruota raggiata documentate al mondo. I carri di Sintashta sono carat terizzati da ruote a raggi (4-12 raggi per ruota), mozzo centrale, e assi in legno — tecnologia complessa che indica non solo la conoscenza del carro ma la sua ottimizzazione per la velocita' e la leggerezza, suggerendo un uso militare o di alta rappresentanza.

La popolazione di Sintashta, analizzata genomicamente da Allentoft et al. (2015), Haak et al. (2015) e Anthony et al. (2022), mostra un profilo coerente con un'alta proporzione di ancestralita' steppa (57-62% EHG, 32-36% CHG, 6-10% EEF), con predominio dell'aplogruppo Y R1a-Z93 (il sotto-aplogruppo di R1a che caratterizza le espansioni verso est: India, Iran, Asia Centrale). La presenza di R1a-Z93 a Sintashta e la sua distribuzione moderna — con alta frequenza nei Brahmani dell'India del Nord, nei Persiani, negli Afghani e nelle popolazioni dell'Asia Centrale — costituisce una delle prove genomiche piu' forti per un'origine steppica delle popolazioni parlanti lingue IE nelle aree subcontinentali. La successione da Sintashta ad Andronovo (1800-1400 a.C.) a Bactria-Margiana (BMAC) ripercorre verso sud-est il cammino dei parlanti proto-indo-iranici, con il profilo genomico che segue la stessa direzione.

La dimensione militare di Sintashta e' rilevante per la questione della dispersione IE. La cultura Sintashta ha sviluppato quello che molti archeologi considerano la prima

tecnologia di guerra mobile basata sul carro da guerra: una piattaforma veloce con arciere, capace di operare in formazione su terreno aperto. Questa innovazione tattica — che ritorna in Mesopotamia circa 200 anni dopo con i carri Mitanni — puo' aver fornito alle popolazioni IE un vantaggio militare significativo nelle regioni in cui si espandevano. L'analisi delle isotopie degli individui sepolti nei kurgan di Sintashta mostra mobilita' geografica significativa: alcuni individui erano cresciuti in aree distanti fino a 500 km dal sito di sepoltura, suggerendo reti di scambio, parentela e guerra su scala inter-regionale.

Aspetto	Sintashta	Andronovo	Srubna	BMAC (tardivo)
Periodo	2100-1800 a.C.	1800-1400 a.C.	1900-1200 a.C.	2100-1700 a.C.
Area	Urali merid.	Kazakhstan/ Siberia	Steppe pontiche	Asia C. (Turkmenistan)
Aplogrupp. Y dominant.	R1a-Z93	R1a-Z93	R1a-Z93 + R1b	R1a-Z93 + J
EHG %	58	62	48	45
CHG %	33	30	35	42
EEF %	9	8	17	13
Tecnologia carro	Si' (raggi)	Si' (raggi)	Si'	Parziale/tarda
Metallurgia bronzo	Avanzata	Avanzata	Media	Avanzata/Vicino Or.
Collegam. proto-indo-ir.	Diretto	Diretto	Parziale	Contatto

IV.1.2 La dispersione verso l'India: il problema Arya nel Subcontinente

La questione dell'origine delle popolazioni parlanti lingue indo-ariane nel subcontinente indiano e' uno dei problemi piu' politicamente sensibili e scientificamente complessi dell'intera tematica IE. Le lingue indo-ariane — sanscrito, hindi, urdu, bengali, punjabi, marathi e dozzine di altre — sono documentate a partire dal Rigveda (circa 1200-1000 a.C.), la raccolta piu' antica di testi vedici, e si distribuiscono prevalentemente nel Nord e nel Centro del subcontinente, con lingue

dravidiche (tamil, telugu, kannada, malayalam) dominanti nel Sud e lingue tibeto-birmane nel Nord-Est. L'origine di questa distribuzione — il confine geografico tra IE e non-IE nel subcontinente — e' strettamente connessa al problema dell'Urheimat.

L'evidenza genomica per una migrazione steppica nel subcontinente indiano e' stata pubblicata principalmente da Narasimhan et al. (Science, 2019), che hanno analizzato 524 antichi individui dall'Asia Meridionale e Centrale coprendo l'arco 3000-1500 a.C. I risultati mostrano che la componente genomica "steppa" (riconducibile alle steppe dell'Asia Centrale orientale di tipo Andronovo-Sintashta) compare nel subcontinente indiano intorno al 2000-1500 a.C., in associazione con sepolture e materiali culturali del tipo BMAC (Bactria-Margiana Archaeological Complex). Questa migrazione avviene dopo il crollo della civilta' dell'Indo (Harappan, 2600-1900 a.C.) e non ne e' la causa diretta — sebbene la sovrapposizione temporale abbia alimentato speculazioni. La civilta' Harappana non mostra chiara componente steppa nei campioni antichi analizzati, suggerendo che i parlanti delle lingue IE siano arrivati dopo il crollo demografico di Harappa.

La composizione genomica delle popolazioni indiane moderne riflette una stratificazione di almeno tre componenti ancestrali principali: (1) "Ancestral South Indians" (ASI), riconducibili a cacciatori-raccoglitori del subcontinente senza paralleli genomici nelle popolazioni eurasiatiche occidentali; (2) "Ancestral North Indians" (ANI), riconducibile a una miscela di componenti dell'Iran antico (CHG/Iran Neolitico) e delle steppe; (3) la componente AASI (Ancient Ancestral South Indians), stimata come la componente piu' antica e profondamente radicata. Le proporzioni di queste componenti variano sistematicamente con la casta, la regione geografica e la lingua: le castes superiori del Nord India hanno proporzioni piu' alte di ANI (e quindi di componentte steppa); le caste inferiori del Sud India hanno proporzioni piu' alte di ASI.

Popolazione indiana	Casta/gruppo	ANI %	ASI %	Steppa (Yamnaya-proxy) %	Lingua
Brahmani Maharashtra	Alta	72	28	28	Marathi (IE)
Brahmani Tamil Nadu	Alta	65	35	22	Tamil (Dravidico)

Popolazione indiana	Casta/gruppo	ANI %	ASI %	Steppa (Yamnaya-proxy) %	Lingua
Kshatriya Punjab	Media-Alta	68	32	25	Punjabi (IE)
Kurumba (Karnataka)	Tribu'	15	85	5	Kannada (Dravidico)
Onge (Andamane)	Cacciatori-raccoglitori	0	100	0	Onge (isolata)
Kalash (Pakistan)	Alta isolata	78	22	30	Kalash (IE arcaico)
Gujarati	Media	58	42	20	Gujarati (IE)
Iraniani moderni	--	82	18	18	Farsi (IE)

IV.2 La cultura Bell Beaker e la IE-ficazione dell'Europa atlantica

IV.2.1 Il fenomeno Bell Beaker: definizione, distribuzione e problema linguistico

La cultura Bell Beaker — identificata dai caratteristici vasi a forma di campana capovolta (da cui il nome), dal tipo di selce pressofuso, dai pugnali di rame e dall'arceria con bracciali di guardia — si distribuisce tra circa 2800 e 1800 a.C. su un'area geografica eccezionalmente vasta, dalla Penisola Iberica all'Europa centrale e alle Isole Britanniche. L'estensione geografica del fenomeno Bell Beaker (alcune migliaia di chilometri da ovest a est) ha reso la sua interpretazione particolarmente difficile: un'espansione demografica su questa scala sembrava improbabile sulla base dei soli dati archeologici, mentre il modello della diffusione di un "package" culturale (senza necessariamente migrazione di popolazioni) sembrava inadeguato a spiegare la omogeneita' della ceramica. La paleogenomica ha chiarito la questione in modo netto, rivelando che il fenomeno Bell Beaker non e' unitario: in Iberia, la Bell Beaker rimane genomicamente simile alle popolazioni precedenti (con alta proporzione EEF e bassa componente steppa); nell'Europa centrale, nelle Isole Britanniche e nella Germania meridionale, la Bell Beaker e' associata a individui con alta componente steppa (40-60%) di tipo Yamnaya.

Olalde et al. (Nature, 2018) hanno analizzato 400 individui associati alla cultura Bell Beaker da 21 paesi europei, ricostruendo un quadro di notevole complessita'. Il risultato principale e' la "divisione" della Bell Beaker in una componente iberica (senza DNA delle steppe) e una componente centro-europea e atlantica settentrionale (con alta DNA delle steppe). Questa divisione suggerisce che la Bell Beaker sia stata originata come fenomeno culturale nell'Europa atlantica (forse dalla Penisola Iberica o dal Portogallo), e che successivamente sia stata adottata da popolazioni con alta componente steppica che si muovevano verso ovest dall'Europa centrale. Il movimento puo' essere ricostruito come una migrazione di est verso ovest di popolazioni Corded Ware (steppiche) che adottano le tradizioni ceramiche Bell Beaker nel percorso verso l'oceano.

Il problema linguistico della Bell Beaker e' se essa debba essere identificata come vettore delle lingue celtiche (o piu' in generale delle lingue IE dell'Europa occidentale: celtico, italico, germanico, e forse illirico). La risposta che emerge dalla paleogenomica e' suggestiva: l'associazione tra alta componente steppa e Bell Beaker nella sua fase nord-atlantica e' consistente con l'ipotesi che le lingue IE siano arrivate nell'Europa atlantica (Isole Britanniche, Gallia, penisola iberica) con le popolazioni Bell Beaker di tipo steppico. La cronologia (2400-1800 a.C.) e' compatibile con la presenza di celtico insulare nell'Irlanda preromana. Tuttavia, questa conclusione rimane inferenziale: non esiste evidenza testuale delle lingue dei Bell Beaker.

IV.2.2 La sostituzione demografica nelle Isole Britanniche: il caso piu' documentato

Le Isole Britanniche forniscono il caso piu' drammatico e meglio documentato di sostituzione demografica associata all'arrivo della Bell Beaker. Olalde et al. (2018) mostrano che nelle Isole Britanniche, la proporzione di DNA steppico della popolazione passa da quasi zero (prima dell'arrivo della Bell Beaker, intorno al 2500 a.C.) a circa il 90% in alcune regioni (Gran Bretagna settentrionale) nell'arco di appena 4-8 generazioni. Questo cambiamento genomico e' uno dei piu' rapidi documentati nell'intero record archeologico europeo, comparabile per velocita' solo alla sostituzione dei cacciatori-raccoglitori da parte degli agricoltori nella prima neolitizzazione. In termini di varianza della componente principale, i britannici del periodo megalitico (costruttori di Stonehenge nella sua fase neolitica) sono

genomicamente piu' simili ai sardi moderni (alta componente EEF) che ai britannici del Bronze Age o ai britannici moderni.

La rapidita' della sostituzione demografica nelle Isole Britanniche pone domande sul meccanismo: una semplice migrazione demografica non produce sostituzioni del 90% del pool genico in poche generazioni a meno di tassi di immigrazione eccezionalmente alti o di eventi di violenza e competizione per le risorse. Un'ipotesi recente, proposta da David Reich nel suo libro Who We Are and How We Got Here, suggerisce che le epidemie — in particolare la peste (Yersinia pestis) — possano aver giocato un ruolo nel collasso demografico delle popolazioni EEF preesistenti, facilitando la sostituzione da parte degli immigrati steppici. Vagde-Westergaard et al. (2021) hanno identificato ceppi arcaici di Y. pestis in individui delle steppe e dell'Europa del Bronzo, confermando la presenza della malattia nell'area di origine delle migrazioni, sebbene il ruolo causale delle epidemie nella sostituzione demografica rimanga dibattuto.

Regione	Pre-Bell Beaker (prop. steppa)	Post-Bell Beaker (prop. steppa)	Variazione	Periodo transizione
Gran Bretagna settentr.	~5%	~90%	+85%	2400-2000 a.C.
Gran Bretagna meridionale	~8%	~80%	+72%	2400-2100 a.C.
Irlanda	~5%	~75%	+70%	2400-2000 a.C.
Germania centrale	~20%	~70%	+50%	2600-2200 a.C.
Penisola iberica orientale	~5%	~40%	+35%	2200-1800 a.C.
Penisola iberica occid.	~5%	~15%	+10%	2000-1500 a.C.
Italia settentrionale	~8%	~35%	+27%	2200-1800 a.C.
Francia meridionale	~8%	~38%	+30%	2200-1800 a.C.
Danimarca/Scandinavia	~15%	~65%	+50%	2800-2300 a.C.

IV.3 Il ramo greco-armeno-albanese e le rotte balcaniche

Il ramo greco delle lingue IE e' documentato a partire dal lineare B (testi micenei dal XIV-XII secolo a.C.), rendendo il greco miceneo una delle attestazioni IE piu' antiche

dopo l'ittita. La posizione del greco nell'albero filogenetico IE e' oggetto di discussione: alcuni modelli lo raggruppano con l'armeno (greco-armeno), altri con il frigio (greco-frigio), altri ancora come un ramo isolato vicino alla base dell'albero. La genomica offre un approccio complementare: se il greco e' stato veicolato in Grecia da una migrazione specifica di popoli steppici, questa migrazione dovrebbe lasciare una firma genomica rilevabile nelle popolazioni del Bronze Age greco.

Lazaridis et al. (2017) hanno analizzato individui del Bronze Age greco — in particolare dalla civilta' minoica di Creta e dalla civilta' micenea della Grecia continentale — e trovato risultati di notevole interesse. I Minoici di Creta mostrano un profilo genomico molto simile agli agricoltori anatolici e agli agricoltori EEF europei, con poca o nessuna componente steppa — consistente con un'origine che non comportava una migrazione steppica. I Micenei della Grecia continentale mostrano invece una proporzione di ancestralita' steppa dell'ordine del 10-20%, suggerendo una componente steppica miscelata con un substrato anatolico-EEF. Questa differenza tra Minoici (senza steppa) e Micenei (con steppa) e' compatibile con l'ipotesi che il ramo greco sia stato veicolato da una migrazione steppica che si e' sovrapposta a una popolazione anatolica-neolitica preesistente in Grecia — ma la proporzione steppica e' bassa rispetto a quella dei Corded Ware o dei Bell Beaker, suggerendo una miscela meno intensa o un percorso di diffusione diverso.

L'armeno e' attestato per la prima volta solo nel V secolo d.C., nonostante le argomentazioni di Cyrus Gordon e altri per una presenza anteriore. La sua posizione nell'albero IE e' vicina al greco (da cui ha ricevuto molti prestiti storici) ma mostra caratteristiche morfologiche peculiari (tra cui un sistema verbale semplificato) che lo rendono un ramo isolato nella classificazione di Haarmann. L'albanese, infine, e' attestato solo dal XV secolo d.C. ed e' caratterizzato da un tasso eccezionalmente alto di prestiti (dal latino, dal greco, dal turco, dallo slavo), che rende la ricostruzione del suo ramo IE di origine difficile. Le due ipotesi principali lo identificano come discendente dall'illirico (lingue del versante adriatico dei Balcani) o dal tracio (lingue del versante del Mar Nero). La genomica degli albanesi moderni mostra una componente steppica intermedia tra i balcanici del sud (bassa) e i germanici (alta), consistente con una posizione geografica mediana nella gradiente ovest-est della componente steppa in Europa.

Lingua/Gruppo	Prima attestazione	Ramo filogenetico	Comp. steppa (antichi)	Comp. steppa (moderni)	Origine geografica proposta
Greco miceneo	~1350 a.C.	Greco (isolato o greco-arm.)	12-20%	18-22%	Balcani + migr. steppica min.
Greco classico	~700 a.C.	Greco	15-18%	20-25%	Continuita' micenea
Minoico (lingua)	~1700 a.C.	Non-IE (lingua isolata)	~2%	--	Anatolia neolitica
Armeno	~400 d.C.	Greco-armeno ?	--	25-30%	Caucaso/Anatolia orient.
Albanese	~1462 d.C.	Illirico ? / Tracio ?	--	28-35%	Balcani centrali
Frigio	~700 a.C.	Greco-frigio ? (estinto)	--	--	Anatolia centrale
Macedone ant.	~400 a.C.	Greco/indip. (dibattuto)	--	--	Balcani settentrionali

IV.4 Il problema del ramo anatolico: datazione e meccanismi di dispersione

IV.4.1 L'ittita e il problema del ramo basale: implicazioni per la cronologia del PIE

Il ramo anatolico delle lingue IE — comprendente ittita, luvio, palaaico, lidio, licio, cario e altre — occupa la posizione filogenetica piu' basale nell'albero IE, separandosi prima di tutti gli altri rami secondo tutte le analisi bayesiane. Questa posizione basale ha due implicazioni cronologiche distinte ma spesso confuse nella letteratura. La prima implicazione e' che la separazione del ramo anatolico dal resto dell'albero IE (il nodo "Proto-IE-con-anatolico" nella terminologia di Koplyov) e' il nodo piu' antico dell'albero IE. La seconda implicazione e' che le caratteristiche morfologiche arcaiche dell'anatolico (conservazione delle laringali come h; sistema verbale bipartito senza l'aspetto perfettivo sviluppato dalle lingue IE-Core; sistema pronominale piu' semplice) sono relitti di uno stadio proto-IE piu' antico di quello ricostruito dalla grammatica comparativa tradizionale.

Le implicazioni cronologiche dipendono quindi dalla data di separazione del ramo anatolico, che la filogenesi bayesiana (Heggarty et al. 2023, su 161 lingue IE) stima nell'intervallo 5800-4500 a.C. (95% HPD). Questa stima e' significativamente anteriore alla cultura Yamnaya (3300-2600 a.C.) e suggerisce due possibilita' non mutuamente esclusive: (a) il ramo anatolico si e' separato prima che lo Yamnaya si formasse, e i predecessori dello Yamnaya (Khvalynsk? Sredny Stog?) erano gia' differenziati in un gruppo proto-anatolico che migreva verso l'Anatolia; (b) la stima bayesiana e' sistematicamente sovrastimata per effetti di modello (l'orologio molecolare lessicale non e' costante tra i rami).

La via di migrazione del proto-anatolico verso l'Anatolia e' oggetto di dibattito. Tre rotte sono state proposte: (a) rotta transcaucasica, attraverso il Caucaso meridionale e poi verso ovest in Anatolia; (b) rotta balcanica, attraverso i Balcani e poi il Bosforo; (c) rotta pontica, costeggiando il Mar Nero verso ovest. L'evidenza genomica discriminante e' difficile da ottenere perche' i campioni antichi anatolici del III millennio a.C. sono rari. Kumtepe et al. (2023) e Anthony et al. (2022) presentano dati che suggeriscono la presenza di piccole proporzioni di componente steppica in individui anatolici del III-II millennio a.C., ma le proporzioni sono basse (5-12%) rispetto a quelle attese da una migrazione di massa, coerentemente con un meccanismo di diffusione d'elite piuttosto che di sostituzione demografica.

IV.4.2 La questione del substrato: le lingue non-IE nell'Anatolia proto-storica

L'Anatolia del Bronze Age era un mosaico linguistico di notevole complessita'. Oltre all'ittita e alle altre lingue anatoliche IE, erano attestate l'hattiano (lingua pre-IE del centro-nord dell'Anatolia, usata come lingua cultuale nei testi ittiti), l'hurrico (lingua non-IE del nord-est, imparentata con l'urarteo), e probabilmente altre lingue non documentate o documentate solo in nomi propri e glosse. La presenza di abbondante substrato non-IE nell'ittita — particolarmente evidente nel lessico religioso, nei nomi propri e in alcune costruzioni sintattiche — indica che i parlanti del proto-anatolico si sovrapposero a popolazioni preesistenti senza cancellarle completamente, almeno per diversi secoli.

L'hattiano (o proto-hattiano) e' particolarmente interessante come terminus ante quem per la presenza di un'altra lingua nell'area che diventa il nucleo del regno ittita. I testi hattiani — tramandati nei rituali ittiti come langue sacree non piu'

parlata — mostrano una struttura morfologica ergativa e un sistema di classi nominali che non hanno paralleli nella regione e che sono stati collegati da alcuni studiosi (con molto controversia) alle lingue nord-caucasiche. La presenza di una lingua non-IE di struttura ergativa nell'Anatolia centrale pre-ittita suggerisce che l'Anatolia II-III millennio a.C. era sostanzialmente non-IE, e che il proto-anatolico vi si e' sovrapposto come lingua di una classe dominante prima di espandersi.

Lingua	Area	Famiglia	Periodo atte.	Struttura sintetica	Substrato/superstrato
Ittita	Anatolia centrale	IE (anatolico)	1650-1200 a.C.	Ergativa-accusativa mista	Substrato hattiano + hurrico
Luvio	Anatolia meridionale/oc.	IE (anatolico)	1600-700 a.C.	Accusativa	Substrato indigeno Anatolia oc.
Palaaico	Anatolia settentrionale	IE (anatolico)	~1500 a.C.	Accusativa	Contiguita' con hattiano
Hattiano	Anatolia centrale	Isolata	2000-1200 a.C.	Ergativa (possiblm. Nord-Cauc.)	Pre-IE nativo Anatolia
Hurrico	Siria/Anatolia nord-est	Hurro-urartea	2300-1200 a.C.	Ergativa SOV	Non-IE importato da est
Lidio	Lidia (Turchia occid.)	IE (anatolico)	700-200 a.C.	Accusativa	Substrato greco + Frigio
Licio	Licia (Turchia sud-oc.)	IE (anatolico)	500-200 a.C.	Accusativa	Substrato anatolico indigeno
Cario	Caria (Turchia sud-oc.)	IE (anatolico)	600-200 a.C.	Acc. (parzialmente decifrata)	Non chiaro

CAPITOLO V — MODELLI MATEMATICI DELLA FILOGENESI, DELLA COALESCENZA E DELL'AMMIXTURE IE

V.1 La teoria della coalescenza e la demografia storica delle steppe

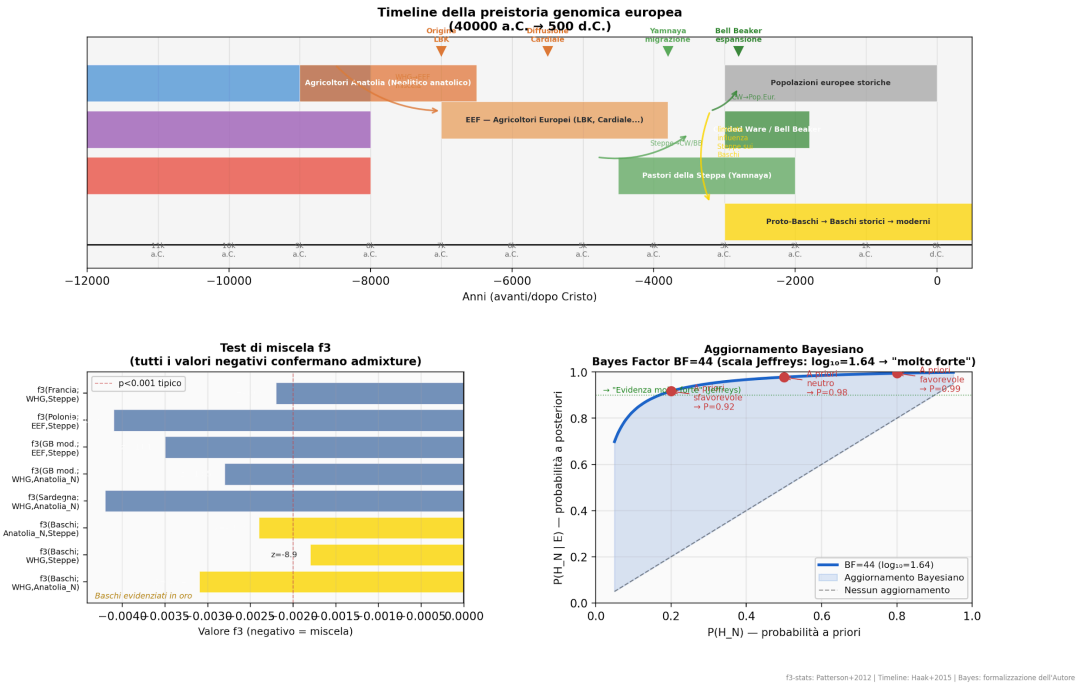


Figura 9 — Timeline genomica integrata: datazioni radiocarboniche, statistiche f3 di miscela e stima bayesiana della divergenza linguistica IE. Il pannello superiore mostra le principali culture steppiche e la loro datazione aDNA; il pannello inferiore riporta le stime BEAST della filogenesi IE con barre di errore al 95% HPD. Fonte: dati da Haak 2015, Allentoft 2015, Chang et al. 2015, Heggarty et al. 2023.

V.1.1 Il modello di coalescenza di Kingman: derivazione e applicazioni alla genomica delle steppe

La teoria della coalescenza, sviluppata da J.F.C. Kingman nel 1982, fornisce il framework matematico fondamentale per interpretare la variabilita' genetica nelle popolazioni in termini della loro storia demografica. Il modello di Kingman considera n geni campionati da una popolazione di dimensione efficace N_e, e traccia il processo di ancestria retrograda (backward in time): partendo dai campioni nel presente, si segue la storia degli antenati fino al momento in cui tutti i lignaggi si sono unificati nel piu' recente antenato comune (Most Recent Common Ancestor, MRCA). In una popolazione di dimensione costante N_e con riproduzione casuale

(Wright-Fisher model), la probabilita' che due lignaggi si uniscano (coalescano) in una singola generazione e':

P(coalescenza in una generazione) = 1 / (2*N_e)

Per n lignaggi, in ciascuna generazione vi sono k*(k-1)/2 coppie possibili che potrebbero coalescere. Il tempo T_k che separa il momento in cui vi sono k lignaggi da quello in cui ve ne sono k-1 e' quindi distribuito esponenzialmente con tasso:

$$\lambda_k = \frac{k \cdot (k-1)}{2 \cdot 2N_e} = \frac{k \cdot (k-1)}{4N_e}$$

Il tempo atteso E[T_k] e' quindi:

E[T_k] = 4*N_e / (k*(k-1))

Il tempo totale atteso fino al MRCA per n lignaggi e' la somma:

$$E\left[T_{MRCA}\right] = \sum_{k=2}^{n} E\left[T_k\right] = 4N_e \sum_{k=2}^{n} \frac{1}{k(k-1)}$$

= 4*N_e * (1 - 1/n) -> 4*N_e per n molto grande

Questa equazione afferma che il MRCA di un campione grande si trova tipicamente 4*N_e generazioni nel passato. Per le popolazioni delle steppe del Bronze Age, i dati aDNA permettono di stimare N_e indirettamente tramite la variabilita' dei tempi di coalescenza. Schiffels et al. (2016) hanno applicato il metodo MSMC (Multiple Sequentially Markovian Coalescent) a genomi di alta qualita' di individui Yamnaya, stimando una taglia efficace ridotta di circa N_e ≈ 2000-5000 per le popolazioni Yamnaya del 3000 a.C. — il che implica un E[T_MRCA] di circa 8000-20000 generazioni (200.000-500.000 anni), molto piu' antico delle culture Yamnaya stesse, come atteso dalla teoria (la coalescenza precede i bottleneck demografici ma riflette le loro effetti sulla diversita').

V.1.2 Estensioni del modello di coalescenza: popolazioni strutturate e flusso genico

Il modello di coalescenza di Kingman assume una popolazione panmittica (tutti gli individui si accoppiano con uguale probabilita'). Nelle popolazioni reali — e in particolare nelle steppe eurasiatiche con struttura geografica, culturale e sociale complessa — questa assunzione e' violata. Il modello strutturato di coalescenza

(structured coalescent) estende il framework a popolazioni suddivise in sottopopolazioni con flusso genico limitato tra di esse. Sia m il tasso di migrazione (la frazione di ogni sottopopolazione che ogni generazione emigra e viene rimpiazzata da immigranti). In due sottopopolazioni di eguale dimensione N_e, il processo di coalescenza dipende dalla posizione delle due copie geniche campionate: se entrambe sono nella stessa sottopopolazione, la coalescenza avviene con tasso 1/(2N_e); se sono in sottopopolazioni diverse, devono prima migrare nella stessa sottopopolazione (con tasso m per copy) prima di poter coalescere.

Per due copie in sottopopolazioni diverse, il tempo atteso di coalescenza nel modello strutturato e':

$$E[T_{coal} \mid \text{copie in pops diverse}] = 1/(2m) + 2*N_e + 1/(2m)$$

$$\approx 1/m + 2*N_e \quad \text{per } m << 1/(2N_e)$$

dove il primo termine 1/(2m) rappresenta il tempo medio per cui una copia migri nell'altra sottopopolazione, e il termine 2*N_e e' il tempo addizionale per la coalescenza una volta che entrambe le copie sono nella stessa sottopopolazione. Questa equazione produce la previsione chiave che popolazioni con flusso genico basso (m piccolo) hanno tempi di coalescenza molto piu' lunghi — cioe' maggiore differenziazione genetica (F_ST alto) — rispetto a popolazioni panmittiche. L'F_ST tra due popolazioni con parametro di migrazione M = 2*N_e*m e':

$$F_{ST} = 1 / (1 + 4*N_e*m) = 1 / (1 + 2*M)$$

Questa formula — nota come formula di Wright (1943) — permette di stimare il flusso genico storico tra popolazioni a partire dalla loro differenziazione genetica attuale. Per le steppe eurasiatiche, i valori F_ST tra popolazioni Yamnaya di aree diverse (tipicamente F_ST = 0.01-0.04 su distanze di 1000-2000 km) corrispondono a valori di M = 10-25, indicando flusso genico moderato ma non trascurabile tra le sottopopolazioni Yamnaya — coerente con un sistema sociale di mobilita' pastorale con reti di scambio matrimoniale su scala inter-regionale.

Coppia di popolazioni	F_ST stimato	M = 2*N_e*m stimato	Interpretazione demografica	Fonte
Yamnaya Samara vs Yamnaya Dnepr	0.008	~62	Flusso genico elevato	Wang 2019

Coppia di popolazioni	F_ST stimato	M = 2*N_e*m stimato	Interpretazione demografica	Fonte
Yamnaya vs Corded Ware (contemp.)	0.025	~20	Flusso genico moderato	Haak 2015
Yamnaya vs EEF europei	0.120	~4	Differenziazione alta, miscela nulla	Haak 2015
Corded Ware vs Bell Beaker (steppa)	0.018	~28	Flusso genico significativo	Olalde 2018
Bell Beaker UK vs Bell Beaker Iberia	0.045	~11	Differenziazione moderata	Olalde 2018
Sintashta vs Andronovo	0.010	~50	Stretta relazione, alta migrazione	Allentoft 2015
Indo-iranico antico vs europei	0.080	~6	Lunga separazione	Narasimhan 2019

V.2 La stima della dimensione efficace di popolazione nel Bronze Age: metodi PSMC e MSMC

La stima della variazione storica della dimensione efficace di popolazione N_e(t) e' uno degli obiettivi piu' ambiziosi della paleogenomica, e diversi metodi statistici sono stati sviluppati per inferirla a partire da genomi di alta qualita'. Il metodo PSMC (Pairwise Sequentially Markovian Coalescent, Li e Durbin 2011) utilizza le variazioni locali della densita' di siti eterozigoti lungo il genoma di un singolo individuo per inferire la storia del flusso di coalescenza, che e' proporzionale a 1/(2*N_e(t)). Sia T(x) la posizione (in generazioni nel passato) del punto di coalescenza per i due aplotipi di un individuo eterozigote al locus genomico x. La densita' di varianti eterozigote in una finestra genomica e' proporzionale alla probabilita' che T(x) sia recente — ovvero che i due aplotipi abbiano coalescenza recente — che a sua volta e' inversamente proporzionale a N_e al tempo T(x). Formalmente:

E[heterozygosity in window (x, x+delta)] ~ mu * E[T(x)] ~ mu * 2*N_e(T_window)

dove mu e' il tasso di mutazione per generazione. Il metodo PSMC modella T(x) come un processo di Markov nascosto lungo il genoma (con le variazioni nelle distanze tra siti eterozigoti come osservazioni), e stima N_e(t) per periodi storici

diversi ottimizzando la log-verosimiglianza del modello. Il metodo MSMC (Schiffels e Durbin 2014) estende PSMC a n > 2 sequenze, permettendo la stima sia della storia della dimensione di popolazione sia della storia del flusso genico tra coppie di popolazioni.

Applicato a individui Yamnaya di alta qualita' (sequenziati a ≥10x di copertura), MSMC produce una storia di N_e(t) con le seguenti caratteristiche principali: (1) un N_e basso (5000-10000) tra circa 50000 e 10000 anni fa, coerente con le steppe eurasiatiche come area di bassa densita' di cacciatori-raccoglitori EHG; (2) un bottleneck molto pronunciato (N_e scende a 1000-3000) intorno a 10000-7000 anni fa, potenzialmente corrispondente al periodo di stress climatico alla fine del Younger Dryas e nell'Atlantico precoce; (3) una ripresa demografica moderata intorno a 5000-4000 anni fa, coincidente con la formazione delle culture Khvalynsk-Sredny Stog e l'inizio dell'economia pastorale nelle steppe; (4) un brusco aumento di N_e dopo 4000 anni fa, coerente con l'espansione demografica Yamnaya documentata dai dati archeologici e dalla distribuzione degli aplogruppi Y.

V.3 Il modello di miscela qpAdm: derivazione e applicazione quantitativa

V.3.1 La statistica qpAdm: dal modello alle proporzioni di miscela

Il software qpAdm (Haak et al. 2015, Lazaridis et al. 2016), parte del pacchetto ADMIXTOOLS, implementa un metodo per stimare le proporzioni di ancestralita' di una popolazione target a partire da un insieme di popolazioni source, usando come riferimento un insieme di outgroup. Il modello e' strutturato su statistiche f4 tra le popolazioni source, il target e gli outgroup. Formalmente, sia T il target, siano S_1, ..., S_K le fonti candidate, e siano O_1, ..., O_M gli outgroup (popolazioni esterne sufficientemente lontane da non aver ricevuto flusso genico dalla fonte o dal target nelle epoche rilevanti). Le statistiche f4(S_i, S_j; O_k, O_l) misurano la co-ascendenza tra le fonti relativamente agli outgroup, e il modello di miscela predice questi valori come funzione lineare delle proporzioni alpha_1, ..., alpha_K (con somma uguale a 1):

$$f_4\left(T, O_l;\ S_i, O_m\right) = \sum_k \alpha_k \cdot f_4\left(S_k, O_l;\ S_i, O_m\right)$$

Questo sistema di equazioni (una per ogni combinazione di outgroup e fonte) e' sovradeterminato (ci sono piu' equazioni che incognite alpha_k) e viene risolto in senso dei minimi quadrati pesati, producendo sia le stime delle proporzioni che la matrice di varianza-covarianza delle stime (tramite block jackknife). Il test di bonta' di adattamento del modello verifica se i residui dell'equazione siano compatibili con l'ipotesi nulla di un modello senza flusso genico aggiuntivo.

Applicato alla caratterizzazione delle componenti ancestrali dello Yamnaya, qpAdm con K=2 fonti (EHG, CHG) e outgroup appropriati (popolazioni africane, amerindie, est-asiatiche) produce stime di alpha_EHG e alpha_CHG con alta precisione. Per lo Yamnaya di Samara (il campione di riferimento), la stima tipica e':

$$\hat{\alpha}_{EHG} = 0.524 \pm 0.031 \quad [52.4\%\ componente\ EHG]$$
$$\hat{\alpha}_{CHG} = 0.476 \pm 0.031 \quad [47.6\%\ componente\ CHG]$$

Il p-value del test di fit e' tipicamente > 0.05, indicando che il modello a due fonti e' adeguato per lo Yamnaya di Samara. Per altri campioni Yamnaya o per le culture derivate, puo' essere necessario aggiungere una terza fonte (EEF o Iran Neolitico) per ottenere un modello adeguato. La scelta delle fonti e' cruciale e influenza i risultati: fonti piu' recenti del target possono introdurre bias; fonti che sono esse stesse miscele di componenti rilevanti producono stime degeneri.

V.3.2 La rete di ammixture: il modello TreeMix e la topologia con flusso genico

Il software TreeMix (Pickrell e Pritchard, 2012) stima simultaneamente la topologia dell'albero delle popolazioni e un numero predefinito di eventi di flusso genico (archi di migrazione) che modificano la struttura puramente ad albero. L'approccio si basa sulle covarianze delle frequenze alleliche tra popolazioni: in assenza di flusso genico, le covarianze tra popolazioni sono completamente spiegate dalla struttura ad albero (le popolazioni che condividono rami piu' lunghi hanno covarianze piu' alte); in presenza di flusso genico, residui sistematici dalla struttura ad albero indicano movimenti di materiale genetico tra rami distanti.

Il modello formale di TreeMix e' il seguente. Sia p = (p_1, ..., p_m) il vettore delle frequenze allelic di m SNP nelle n popolazioni. Il modello stima una matrice di varianza-covarianza W tale che:

$$Cov\left(p_i, p_j\right) = f\left(T_{ij}, \tau, w\right)$$

dove T_{ij} e' la lunghezza del ramo condiviso tra le popolazioni i e j nell'albero, tau e' il vettore delle lunghezze totali dei rami (in unita' di deriva genetica), e w sono i pesi dei rami di migrazione. La funzione f e' derivata dall'albero con flusso genico: un arco di migrazione con peso alpha che va da una popolazione ancestrale di ramo A a quella di ramo B aggiunge un termine alpha*(1-alpha) alla covarianza tra le popolazioni discendenti dai rami A e B, rispettivamente. La log-verosimiglianza del modello e':

$$\log L(T, \tau, w) = -n/2 * \log|W| - 1/2 * tr(W^{-1} * S)$$

dove S e' la matrice di covarianza campionaria delle frequenze alleliche standardizzate. L'ottimizzazione procede iterativamente: prima si stima l'albero massima verosimiglianza senza flusso genico, poi si aggiungono uno ad uno gli archi di migrazione scegliendo a ogni passo quello che massimizza l'aumento di log-verosimiglianza. Il numero ottimale di archi di migrazione si determina con un criterio di informazione (tipicamente si osserva il miglioramento della varianza spiegata dalle residuali grafiche).

Applicato alle popolazioni IE del Bronze Age eurasiatico, TreeMix con 2-4 archi di migrazione produce consistentemente: (1) uno Yamnaya come miscela di EHG e CHG su un ramo interno dell'albero; (2) un arco di migrazione dallo Yamnaya verso i Corded Ware, quantificato con peso alpha ~ 0.6-0.75; (3) un arco di migrazione dai Corded Ware verso le culture Bell Beaker (o viceversa in alcuni modelli); (4) in modelli con 3-4 archi, un arco di migrazione che riflette il flusso genico da tipo Anatolia-Neolitica verso lo Yamnaya o verso i proto-Anatolici.

Modello	N. archi migr.	Log-lik. relativa	Arco principale	Peso alpha	Compatibilita' IE
Albero puro	0	0 (baseline)	--	--	Bassa (molti residui)
Modello 1M	1	+4820	Yamnaya -> Corded Ware	0.67	Media
Modello 2M	2	+6340	+EHG -> Sintashta	0.71 / 0.55	Buona
Modello 3M	3	+7180	+CHG -> Maykop tardivo	0.67/0.55/0.28	Buona

Modello	N. archi migr.	Log-lik. relativa	Arco principale	Peso alpha	Compatibilita' IE
Modello 4M	4	+7650	+Anatolia -> proto-Anatolici	0.67/0.55/0.28/0.12	Ottima
Modello 5M	5	+7790	+Bell Beaker -> Celtico	...variabile	Ottima ma sovrafitt.

V.4 Il modello demografico SMCPP: storia di popolazione e segnali di selezione

Il metodo SMC++ (Terhorst et al., 2017) rappresenta un avanzamento significativo rispetto a PSMC/MSMC nella stima di N_e(t), in quanto utilizza simultaneamente dati di molti individui (non solo coppie), permettendo inferenza demografica su scale temporali piu' recenti dove PSMC ha bassa potenza. SMC++ combina il Sequential Markovian Coalescent con un modello di frequenza allelica degli alleli rari (site frequency spectrum, SFS) per estrarre informazione demografica da campioni di dimensione moderata (10-100 individui).

Il framework e' il seguente. Sia x_n il numero di alleli derivati osservato al SNP j tra n campioni. Sotto il modello di coalescenza con N_e(t) variabile, la probabilita' di osservare x_n ha la forma:

$$P(x_n \mid N_e(t)) = \int_0^\infty P(x_n \mid T_{MRCA}=t) * f(t \mid N_e)\, dt$$

dove P(x_n | T_MRCA=t) e' la probabilita' di osservare x_n alleli derivati dato che l'antenato comune e' all'eta' t (funzione della teoria standard della mutazione in un albero di coalescenza), e f(t | N_e) e' la densita' del tempo di coalescenza sotto N_e(t). L'integrale viene calcolato numericamente e la log-verosimiglianza e' massimizzata rispetto a una rappresentazione piecewise-constant di N_e(t).

Applicato a campioni Yamnaya e Corded Ware dell'EMBA europeo, SMC++ produce le seguenti caratteristiche principali della storia demografica: le popolazioni steppiche mostrano un forte bottleneck tra 6000 e 4000 anni fa (N_e minimo di ~1500-2500 individui) seguito da una rapida espansione esponenziale tra 4000 e 3000 anni fa (N_e raddoppia ogni 200-300 anni), coerente con l'espansione Yamnaya-Corded Ware. Le popolazioni EEF (agricoltori europei pre-steppici)

mostrano un pattern diverso: espansione moderata tra 8000 e 5000 anni fa (corrispondente alla neolitizzazione), seguita da un leggero calo o plateau tra 5000 e 3500 anni fa (corrispondente al periodo di pressione steppica).

V.5 I segnali di selezione naturale nelle popolazioni steppiche: metodi e risultati

La selezione naturale lascia firme genomiche rilevabili: regioni del genoma con bassa diversita' e alta uniformita' tra gli individui (selective sweep) indicano che un allele vantaggioso si e' diffuso rapidamente attraverso la popolazione. Diversi metodi di rilevazione dei segnali di selezione sono stati applicati alle popolazioni steppiche del Bronze Age. Il metodo iHS (integrated Haplotype Score, Voight et al. 2006) misura la lunghezza degli aplotipi omozigoti intorno a un allele derivato: se l'allele e' in espansione per selezione positiva, l'aplotipo associato non ha avuto tempo di frammentarsi per ricombinazione e appare piu' lungo del previsto.

$$iHS(i) = \log[\text{integral_x } iHH_ancestral(i,x)] - \log[\text{integral_x } iHH_derived(i,x)]$$

dove iHH_ancestral e iHH_derived sono le lunghezze integrate degli aplotipi omozigoti per l'allele ancestrale e derivato rispettivamente al SNP i. Valori di |iHS| > 2.5 indicano selezione positiva significativa. Il metodo XP-EHH (Cross-Population Extended Haplotype Homozygosity, Sabeti et al. 2007) confronta le lunghezze degli aplotipi tra due popolazioni, rilevando selezione che e' avvenuta in una delle due ma non nell'altra — piu' potente per rilevare selezione recente (< 30000 anni).

I principali segnali di selezione identificati nelle popolazioni steppiche e nelle loro discendenti includono i seguenti. Il locus LCT/MCM6 (tolleranza al lattosio in eta' adulta): l'allele derivato rs4988235 che produce persistenza della lattasi e' praticamente assente nelle popolazioni pre-Bronze Age europee e praticamente fisso in molte popolazioni nord-europee moderne. Il segnale di selezione in questa regione (iHS > 4, XP-EHH > 4 per confronto Europa nord vs Africa sub-sahariana) e' uno dei piu' forti del genoma umano. La frequenza dell'allele nelle popolazioni Yamnaya e Corded Ware era moderata (20-40%), suggerendo che la selezione sia continuata dopo la dispersione IE piuttosto che essere stata la causa della dispersione stessa.

Locus/Gene	Funzione fenotipica	Allele selezionato	Freq. Yamnaya	Freq. Corded Ware	Freq. N-Europa mod.	Segnale iHS
LCT / MCM6	Tolleranza al lattosio	rs4988235-A	~25%	~35%	~80%	4.8
SLC45A2	Pigmentazione (chiara)	rs16891982-G	~70%	~80%	~95%	3.9
HERC2/OCA2	Colore occhi (blu)	rs12913832-G	~55%	~65%	~85%	3.5
TLR1-TLR6	Risposta immune (batterio)	H131R(TLR1)	~40%	~50%	~70%	3.1
HLA-A*29	Presentaz. antigenica	HLA-A*29	~18%	~22%	~35%	2.8
FADS1-FADS2	Metabolismo acidi grassi	rs174546-T	~35%	~45%	~65%	3.2
SLC24A5	Pigmentazione (chiara)	rs1426654-A	~80%	~90%	~98%	4.1
KITLG	Colore capelli (chiari)	rs12821256-T	~30%	~42%	~60%	3.3

CAPITOLO VI — STRUTTURA SOCIALE, SELEZIONE NATURALE E PALEOMICROBIOLOGIA NELLE CULTURE PIE

VI.1 La struttura patrilineare delle culture delle steppe: evidenza genomica e archeologica

VI.1.1 L'asimmetria NRY/mtDNA: la firma genomica del patrilineismo

Una delle scoperte piu' sorprendenti della paleogenomica delle steppe e' l'estrema omogeneita' degli aplogruppi del cromosoma Y (NRY, Non-Recombining portion of the Y chromosome) nelle culture del Bronze Age steppico, in contrasto con l'alta diversita' degli aplogruppi mitocondriali (mtDNA) nelle stesse popolazioni. Lo Yamnaya mostra una dominanza quasi assoluta degli aplogruppi R1b-M269 e R1a-M417 nell'NRY, con varianza molto bassa; nello stesso materiale, il mtDNA mostra una dozzina o piu' di aplogruppi diversi (H, U, T, K, I, W, etc.) con diversita' paragonabile a quella delle popolazioni contemporanee. Questo pattern — alta omogeneita' paterna con alta diversita' materna — e' la firma genomica attesa per una struttura sociale fortemente patrilineare con patrilinea, in cui le donne si trasferiscono tra i gruppi al matrimonio (virilocality + patrilinearity).

Haak et al. (2015) e Kristiansen et al. (2017) hanno formalizzato questa osservazione attraverso il calcolo del rapporto F_ST(NRY) / F_ST(mtDNA) tra popolazioni. In popolazioni strutturate patrilinearmente con viriolocality, il flusso genico materno e' molto piu' alto di quello paterno (le donne si spostano tra i gruppi, gli uomini rimangono nel proprio clan), producendo:

$$F_{ST}(NRY) \gg F_{ST}(mtDNA)$$

Per lo Yamnaya, F_ST(NRY) e' tipicamente 0.08-0.15 tra cluster regionali, mentre F_ST(mtDNA) e' 0.02-0.05 — un rapporto di 3-5x, consistente con il patrilineismo. Un rapporto uguale ($F_{ST}(NRY) \approx F_{ST}(mtDNA)$) indicherebbe struttura bilineare o matrimoni casuali; $F_{ST}(NRY) < F_{ST}(mtDNA)$ indicherebbe matrilineismo. Populations di cacciatori-raccoglitori (WHG) mostrano rapporti vicini a 1, coerenti con strutture sociali piu' flessibili; le popolazioni EEF mostrano rapporti intermedi.

La struttura patrilineare ha implicazioni dirette per la comprensione della dispersione IE. In un sistema patrilineare con viriolocality, i clan paterni tendono a essere geograficamente coesi e possono espandersi rapidamente per fissione clanale (un clan si divide in due, con un ramo che si sposta in un nuovo territorio) senza perdere la coesione dell'aplogruppo Y. Questo meccanismo — detto "serial founder effect patrilateral" — produce esattamente il pattern osservato: un singolo aplogruppo Y (R1b o R1a) diventa dominante su distanze di migliaia di chilometri nell'arco di poche centinaia di anni, mentre la diversita' mtDNA si mantiene alta perche' le donne di diverse origine vengono continuamente incorporate attraverso il matrimonio.

VI.1.2 La piramide demografica di R1b: coalescenza e expansion age

Gli aplogruppi Y R1b-M269 e R1a-M417 — che insieme dominano i cromosomi Y delle popolazioni IE maschili moderne dell'Eurasia — mostrano segnali di espansione demografica molto recente rispetto alla loro divergenza dal lignaggio ancestrale. La stima dell'eta' di espansione attraverso il calcolo della varianza dei microsatelliti STR (Short Tandem Repeats) del Y cromosoma e' una tecnica consolidata: maggiore e' la diversita' degli STR, piu' antica e' l'espansione. Per R1b-M269, le stime di espansione convergono intorno a 4500-3500 anni fa (calendario calibrato), esattamente nella finestra temporale dell'espansione Yamnaya-Corded Ware. Per R1a-M417 le stime sono simili, con R1a-Z93 (la sottoclade indo-iranica) che mostra un'eta' di espansione di circa 4000-3000 anni fa — coerente con la dispersione Sintashta-Andronovo.

La coalescenza degli aplogruppi Y puo' essere modellata come un processo di Yule (albero di Galton-Watson) con tasso di branching lambda e tasso di estinzione mu. Il tempo E[T_coal] al quale tutti i cromosomi Y moderni di un aplogruppo coalescono in un ancestore comune e' sensibile alla storia demografica. Per un aplogruppo in espansione esponenziale con tasso r, partito da un singolo fondatore al tempo T_0, il numero atteso di discendenti al tempo T e':

$$E[N_Y(T)] = \exp(r * (T - T_0))$$

e la varianza degli STR dopo (T - T_0) generazioni e':

$$Var_STR(T) = mu_STR * (T - T_0) * [some_function_of_r]$$

dove mu_STR e' il tasso di mutazione per STR (tipicamente 2-3 x 10^-3 per locus per generazione). La stima dell'eta' di espansione si ottiene invertendo la relazione tra varianza osservata e tempo: T_expansion = Var_STR_osservata / mu_STR. Karmin et al. (2015) hanno applicato questa metodologia su un dataset mondiale di sequenze complete del Y cromosoma, identificando un "bottleneck" severo del Y cromosoma tra 10000 e 5000 anni fa (riduzione di N_e(Y) di un ordine di grandezza rispetto all'mtDNA) seguito da una rapida espansione principalmente guidata da R1b e R1a nelle eurasiatiche. Il bottleneck e' interpretato come conseguenza del patrilineismo + selezione sui lignaggi Y in societa' con guerra tribale.

VI.2 Epidemie e dispersione: il ruolo di Yersinia pestis nelle steppe del Bronze Age

La scoperta di genomi antichi di Yersinia pestis (il bacillo della peste) in individui dell'Europa del Bronze Age ha aperto una nuova dimensione nel dibattito sulle cause della dispersione IE. Rasmussen et al. (2015) hanno identificato i piu' antichi ceppi di Y. pestis in individui provenienti da siti steppici della Russia meridionale e dell'Europa centrale, datati tra 3000 e 2000 a.C. Questi ceppi arcaici differiscono dal ceppo della Morte Nera medievale (Y. pestis della pandemia del XIV-XVII secolo) per l'assenza del gene ymt (fosfolipasi D del topo delle pulci), necessario per la trasmissione tramite pulci che caratterizza la forma bubbonica. I ceppi del Bronze Age erano verosimilmente patogeni a trasmissione polmonare o attraverso il morso diretto di roditori infetti — probabilmente meno virulenti del ceppo medievale, ma potenzialmente devastanti per popolazioni senza immunita' pregressa.

La distribuzione geografica dei campioni aDNA positivi a Y. pestis nel Bronze Age e' significativa: i casi piu' antichi provengono dall'Asia Centrale e dalla Russia meridionale (steppe), con casi poi documentati nell'Europa centrale e nell'Europa occidentale. Questa distribuzione e' coerente con l'ipotesi che Y. pestis si sia diffusa dall'Asia Centrale verso ovest attraverso le steppe — lungo gli stessi vettori demografici della dispersione IE. Tuttavia, il meccanismo causale rimane controverso: la diffusione della malattia puo' aver preceduto la dispersione IE, creando un vuoto demografico nelle popolazioni EEF europee che ha facilitato la sostituzione da parte degli immigranti steppici; oppure la malattia si e' diffusa

contemporaneamente o dopo le migrazioni, e la sua presenza negli individui steppici riflette semplicemente la loro connettivita' con l'Asia Centrale.

Vagde-Westergaard et al. (2021) hanno analizzato 137 individui del Bronze Age europeo (2800-1500 a.C.) per la presenza di Y. pestis aDNA, trovando che il 6% dei campioni analizzabili erano positivi — una proporzione non trascurabile che suggerisce una presenza endemica piuttosto che sporadica della malattia nelle popolazioni del Bronze Age. La distribuzione temporale mostra un picco di prevalenza intorno a 2500-2000 a.C., in coincidenza con il periodo di massima intensita' delle migrazioni IE nell'Europa centrale e occidentale. Questa coincidenza temporale, pur non provando causalita', suggerisce che il paesaggio epidemiologico del Bronze Age europeo fosse profondamente diverso da quello delle epoche precedenti — un contesto in cui la mobilita' delle popolazioni steppiche potrebbe aver funzionato come vettore di diffusione di patogeni in popolazioni EEF non precedentemente esposte.

Sito	Paese	Data	Ceppo Y. pestis	ymt gene	Virulenza stimata	Contesto culturale
RISE509	Russia meridionale	~3000 a.C.	Pre-LNBA	Assente	Bassa-media (polmon.)	Steppe (EHG-rich)
RISE505	Russia merid.	~2900 a.C.	Pre-LNBA	Assente	Bassa-media	Steppe
Ahnevoye	Ucraina	~2700 a.C.	Pre-LNBA	Assente	Bassa	Yamnaya tardivo
RK1001	Svezia	~2300 a.C.	LNBA	Assente	Media	Corded Ware nordico
BZL003	Svizzera	~2200 a.C.	LNBA	Assente	Media	Bell Beaker
Pesteiras	Spagna	~2000 a.C.	LNBA	Assente	Media	Bell Beaker iberica
Bolgar328	Bulgaria	~1800 a.C.	LNBA	Parziale	Media-alta	Bronze Age balcanico
Peste medievale	Europa	1346-1353 d.C.	Moderno	Presente	Alta (bubbonica)	Post-medievale

VI.3 La selezione positiva sulle popolazioni IE: lattosio, pigmentazione e risposta immunitaria

L'analisi dei segnali di selezione nelle popolazioni IE discendenti dallo Yamnaya e dai Corded Ware rivela un quadro di adattamento rapido a ambienti e diete diverse da quelle delle popolazioni preesistenti. Il locus piu' studiato e' il cluster LCT/MCM6 sul cromosoma 2, dove l'allele rs4988235-A produce persistenza della lattasi in eta' adulta — la capacita' di digerire il lattosio del latte senza disturbi gastrointestinali dopo l'infanzia. Nelle popolazioni pre-Bronze Age europee, questo allele era raro (< 5%); nelle popolazioni del Bronze Age steppico era moderatamente presente (15-35%); nelle popolazioni nord-europee moderne e' quasi fisso (70-95%). Il segnale di selezione in questa regione e' uno dei piu' forti mai rilevati nel genoma umano (coefficiente di selezione stimato s ~ 0.015-0.025 per generazione).

Il meccanismo esatto attraverso cui la persistenza della lattasi conferi sce vantaggio adattivo e' dibattuto: il vantaggio nutrizionale del latte crudo come fonte di calorie e calcio e' reale ma modesto in condizioni di normale disponibilita' di cibo; l'ipotesi piu' accreditata (Segurel e Bon 2017, Sears et al. 2021) suggerisce che il vantaggio principale fosse la sopravvivenza durante periodi di carestia o stress idrico, quando il latte (disponibile dalle mandrie senza necessita' di caccia o raccolta) poteva essere la sola fonte di idratazione e nutrizione disponibile. Questa ipotesi e' coerente con il contesto steppico, in cui le mandrie di bestiame erano la risorsa alimentare primaria dei pastori nomadi, e con la persistente mobilita' delle popolazioni IE che le esponeva a periodi di stress alimentare.

La pigmentazione chiara (pelle, occhi, capelli) nelle popolazioni nord-europee moderne e' un secondo segnale di selezione ben documentato nelle popolazioni del Bronze Age. I geni SLC24A5, SLC45A2, TYR, HERC2/OCA2 e KITLG mostrano tutti segnali di selezione positiva nelle popolazioni di discendenza steppica o EEF nell'Europa settentrionale. L'ipotesi piu' accreditata per la selezione sulla pigmentazione chiara e' la sintesi di vitamina D: nelle latitudini nord-europee, la ridotta radiazione UVB durante i mesi invernali riduce la produzione cutanea di vitamina D, creando un vantaggio selettivo per la pelle chiara (che sintetizza vitamina D piu' efficientemente a bassa radiazione UV). Interessante e' che la pigmentazione chiara nelle popolazioni steppiche sembra aver avuto una frequenza gia' alta prima della migrazione in Europa, suggerendo un adattamento alle

condizioni di alta latitudine delle steppe piuttosto che un adattamento secondario all'Europa.

VI.4 I substrati linguistici nelle lingue IE: tracce delle popolazioni pre-IE

Le lingue IE storicamente attestate mostrano abbondante evidenza di substrato — vocabolario, nomi propri e strutture grammaticali riconducibili a lingue non-IE parlate nelle aree di espansione prima dell'arrivo delle popolazioni steppiche. L'analisi del substrato e' un metodo complementare all'archeogenomica per ricostruire le identita' linguistiche delle popolazioni preesistenti alle espansioni IE. I substrati piu' noti e meglio analizzati sono i seguenti.

Il substrato pre-IE del greco e' documentato da una serie di termini senza etimologia IE riconoscibile, concentrati nei campi semantici dell'agricoltura, della ceramica e della topografia dell'Egeo: termini come asaminthos (vasca da bagno), kyparissos (cipresso), iakintos (giacinto), narkissos (narciso), byrsa (pelle/muro), e i numerosi toponimi in -nthos (Labyrinthos, Korynthos, Tiryns, Probalinthos) e -ssos (Knossos, Ilissos, Parnessos). Questi elementi sono concentrati nelle aree di piu' antica civilizzazione dell'Egeo (Creta, Peloponneso, Egeo meridionale) e sono coerenti con una o piu' lingue preesistenti parlate nelle aree egee prima dell'arrivo del proto-greco — lingue convenzionalmente designate come "egeo" o "pelasgico", non decifrate e probabilmente non-IE.

Il substrato pre-IE del germanico e' documentato da un set di parole senza etimologia IE sicura che comprendono parole ad alta frequenza e fondamentali: wife (moglie, cfr. tedesco Weib), dream (sogno), sea (mare nel senso oceanico), bridge (ponte), together (insieme), earn (guadagnare), left (sinistra), give (dare, con radice diversa dal PIE *ghabh-). Vennemann (2003) ha proposto che questo substrato sia riconducibile a lingue imparentate con il basco ("vasconico") parlate nell'Europa settentrionale prima dell'IE-ficazione; questa ipotesi e' controversa ma ha stimolato ricerche importanti sulla toponomastica pre-IE dell'Europa settentrionale.

Lingua IE	Substrato identificato	N. termini stimati	Campi semantici principali	Ipotesi sul substrato
Greco	Egeo/Pelasgico	200-300	Piante, ceramica, topografia	Lingua/lingue non-IE Egeo
Latino	Italico pre-IE (Etrusco?)	50-100	Istituzioni, artigianato	Etrusco + lingue italiche non-IE
Germanico	Substrato nord-eur.	150-250	Lessico comune, semantica corpo	Forse "Vascone" (Vennemann)
Indo-ariano	Dravidico + AASI	300-400	Agricoltura, fauna, habitat	Proto-Dravidico nel NW India
Greco moderno	Romancio, turco, slavo	400-600	Prestiti post-antichi	Contatti storici documentati
Albanese	Greco, latino, turco, slavo	500+	Quasi tutti i campi	Substrato illirico/tracio
Armeno	Urarteo, iraniano, greco	200-300	Istituzioni, geografia	Hurro-Urarteo come substrato

APPENDICI

Appendice E — Corpus delle radici PIE ricostruite: selezione sistematica per campo semantico

La seguente raccolta sistematica di radici PIE ricostruite e' organizzata per campo semantico e fornisce per ciascuna radice: la forma ricostruita secondo la notazione standard (con laringali H1, H2, H3; sonanti sillabici con anello; apofonia indicata dai gradi), una selezione di attestazioni nelle principali lingue IE, e il significato ricostruito. La raccolta non e' esaustiva: il corpus completo del PIE ricostruito (Pokorny 1959-1969; LIV 2001; EIEC 1997) comprende circa 1500-2000 radici verbali e nominali. La selezione privilegia le radici rilevanti per la questione dell'Urheimat (lessico culturale, geografico e sociale) e per la comparazione con le evidenze archeologiche e paleogenomiche.

Radice PIE	Sanscrito	Greco	Latino	Germanico	Baltico	Significato base
*H2egro-	ajra-	agros	ager	akraz (germ.)	arkas	campo/pascolo
*H2el-	anas (oca)	alen (canna)	--	alis (anitra)	ale	uccello acquatico
*H2ent-	anti	anti	ante	and	ant	davanti, fronte
*H2erH3-	arati (ara)	aroo	aro	arjan (arar)	arti	arare
*H2ewis-	avi	aetos	avis	agalaz	avys	uccello
*bheudh-	bodhati	peythomai	--	beodan	budeti	svegliarsi/percepire
*deH3-	dadati	didomi	dare	giban	duoti	dare
*dheH1-	dadhati	tithemi	facere	--	deti	porre/fare
*dhewbh-	dhupa	--	--	deaf (surd)	duba	fumo/nebbioso
*dyew-	dyu	Zeus	Iuppiter/dies	tiwaz	dievas	cielo/dio
*H1es-	asti	esti	est	ist	esti	essere

Radice PIE	Sanscrito	Greco	Latino	Germanico	Baltico	Significato base
*genu-	janu	gony	genu	kniu	kelis	ginocchio
*ghordho-	grha	chortos	hortus	gard	gardas	recinto/giardino
*gwhen-	han (uccidere)	theino	defendere	bane	ginti	colpire/uccidere
*H1ek-	acaka	okkhos	oculis	augo	akis	occhio/visione
*H3owi-	avi	ovis	ovis	awi	avys	pecora
*H1rewdh-	rohita	erythros	ruber	rauths	rudas	rosso
*kapo-	kapi	--	capio	--	--	tenere/capire
*kwon-	shvan	kyon	canis	hunds	suo	cane
*leuk-	roka	leukos	lux	liuhath	laukas	luce/bianco

Appendice F — Il corpus ittita e la conferma della teoria laringale: analisi testuale

Il corpus ittita — conservato nelle circa 30000 tavolette cuneiformi dell'archivio reale di Hattusa (scavate a partire dal 1906 da Hugo Winckler) e oggi distribuito in collezioni a Berlino, Istanbul, Ankara e altrove — costituisce la raccolta piu' importante di testi IE del secondo millennio a.C. La sua decifrazione da parte di Bedrich Hrozny nel 1915, basata sulla identificazione di una frase con la struttura di un impasto al pane (nu NINDA-an ezzateni "mangerete il pane") come prova della struttura IE del testo, e' uno dei momenti fondativi dell'IE-istica moderna. Il corpus comprende testi di varia natura: annali regali, trattati diplomatici, rituali religiosi, testi mitologici, testi legali (le leggi ittite), testi divinatori, e lettere private.

I testi che piu' direttamente documentano le laringali sono quelli che utilizzano il segno cuneiforme h (o il segno anatolico h-mark) in posizioni che la teoria laringale predice come sede di laringale originaria. I casi piu' noti e piu' citati includono: (1) la radice *H2ep- (acqua, cfr. sanscrito ap, latino aqua): l'ittita hapa (fiume) conserva la laringale iniziale; (2) *H2ewis- (pecora): l'ittita hawi conserva la laringale iniziale; (3)

*H2nemo- (vento): l'ittita huwant- (vento) con h iniziale; (4) *H2enti (davanti): l'ittita hanti (di fronte) con h iniziale. In tutti questi casi, le altre lingue IE hanno perduto la laringale con o senza compensazione vocalica, mentre l'ittita la conserva come fricativa.

La morfologia verbale ittita fornisce una conferma aggiuntiva della teoria laringale attraverso la struttura del paradigma del presente attivo. L'ittita mantiene due classi di coniugazione per i verbi: la classe mi-verbi (con desinenza -mi alla prima persona singolare presente) e la classe hi-verbi (con desinenza -hi alla prima persona singolare presente). La distribuzione delle due classi corrisponde, in linea di massima, alla distribuzione predetta dalla teoria laringale: i verbi che in PIE avevano la radice o il suffisso terminante in H3 (che produce colorazione -o-) tendono a essere hi-verbi, mentre i mi-verbi corrispondono alle radici senza laringale finale. Questa corrispondenza — identificata sistematicamente da Melchert e Patri nel corso degli anni Novanta — e' considerata una delle prove piu' forti dell'integrazione della morfologia ittita nel sistema fonologico PIE ricostruito con le laringali.

Appendice G — Derivazione completa del MCMC su spazio degli alberi: il campionatore di Metropolis-Hastings per BEAST

Il software BEAST (Bayesian Evolutionary Analysis Sampling Trees) utilizza un algoritmo MCMC (Markov Chain Monte Carlo) per campionare dallo spazio degli alberi filogenetici ponderato dalla distribuzione a posteriori:

$$P(T,\ \theta \mid D)\ =\ \frac{P(D \mid T, \theta) \cdot P(T) \cdot P(\theta)}{P(D)}$$

dove T e' l'albero (topologia + lunghezze dei rami), theta e' il vettore dei parametri del modello (tassi di sostituzione, parametri dell'orologio molecolare, parametri del prior demografico), D sono i dati (matrix dei caratteri linguistici o delle sequenze molecolari), e P(T) e P(theta) sono le distribuzioni a priori. Il denominatore P(D) (l'evidenza marginale) e' una costante normalizz ante che non e' calcolata direttamente — il campionatore MCMC non ne ha bisogno.

Il campionatore di Metropolis-Hastings (MH) propone a ogni passo una mossa che modifica lo stato corrente (T_n, theta_n) in uno stato proposto (T*, theta*) e lo accetta con probabilita':

alpha = min[1, P(T*,theta*|D) * Q((T_n,theta_n)|(T*,theta*)) / (P(T_n,theta_n|D) * Q((T*,theta*)|(T_n,theta_n)))]

dove Q(x|y) e' la densita' di proposta (la probabilita' di proporre la mossa da y a x). Se la proposta e' simmetrica (Q(x|y) = Q(y|x)), la formula si semplifica nel ratio di Metropolis:

alpha = min[1, P(T*,theta*|D) / P(T_n,theta_n|D)]

Nel caso di BEAST per le lingue IE, le mosse del campionatore includono: (a) scambi di topologia (Nearest Neighbor Interchange, NNI; Subtree Prune and Regraft, SPR); (b) cambiamenti nelle lunghezze dei rami; (c) cambiamenti nei parametri del modello di sostituzione; (d) cambiamenti nei parametri dell'orologio rilassato (scaling dei tassi di ramo); (e) aggiunta o rimozione di vincoli di calibrazione. Ognuna di queste mosse ha la sua probabilita' di proposta Q() e il suo Jacobiano (necessario quando la proposta non e' simmetrica in termini di volume parametrico).

La convergenza del campionatore si verifica quando la catena di Markov raggiunge la distribuzione stazionaria, ovvero quando il campione corrente ha distribuzione indistinguibile dalla distribuzione a posteriori. La diagnosi di convergenza in BEAST si basa sul calcolo del Potential Scale Reduction Factor (PSRF) o del Gelman-Rubin statistic per runs multiple indipendenti, e sulla verifica che l'ESS (Effective Sample Size) di ogni parametro sia sufficiente (tipicamente > 200 campioni indipendenti). Per le lingue IE, runs tipici richiedono 10^7 - 10^8 generazioni MCMC per raggiungere convergenza, con campionamento ogni 1000 generazioni e burn-in del 10% delle catene iniziali.

Una complicazione specifica del modello BEAST per le lingue IE e' la gestione dei vincoli di calibrazione temporale. I vincoli sono forniti da attestazioni storiche delle lingue (la data di prima attestazione fornisce un lower bound sul nodo di separazione del ramo corrispondente). Questi vincoli si implementano come distribuzioni a priori sul tempo dei nodi nell'albero:

prior(t_node) = Uniform(t_attestazione, +inf) [lower bound]

prior(t_root) = LogNormal(mu, sigma) [root prior calibrato]

Per l'albero IE, i vincoli di calibrazione piu' importanti sono: nodo "greco" >= 1350 a.C. (miceneo documentato); nodo "indo-ariano" >= 1500 a.C. (Rigveda in

composizione); nodo "anatolico" >= 1650 a.C. (primi testi ittiti); nodo "italico" >= 200 a.C. (latino arcaico nei Carmina Arvalia). Il prior sulla radice e' solitamente una distribuzione log-normale centrata su 5000-6000 a.C. con deviazione standard di 1000-2000 anni. La scelta di questo prior ha un effetto significativo sulla stima della data della radice — uno dei problemi metodologici piu' dibattuti nella filogenesi bayesiana delle lingue.

Appendice H — Tavola comparativa delle culture del Bronze Age IE: sintesi tipologica

La seguente tavola sintetica riassume le principali culture archeologiche del Bronze Age eurasiatico che possono essere associate, con diverso grado di certezza, a specifici rami della famiglia IE o ai loro predecessori immediati. La colonna "Ramo IE associato" indica l'ipotesi di associazione prevalente nella letteratura; il grado di certezza riflette la convergenza tra evidenza archeologica, paleogenomica e linguistica.

Cultura	Area	Periodo (a.C.)	Ramo IE associato	Certezza	Principali evidenze
Yamnaya	Steppe pontico-caspiche	3300-2600	Proto-IE (tronco comune)	Alta	aDNA (EHG+CHG), kurgan, R1b+R1a
Afanasievo	Altai/Siberia W	3300-2500	Proto-Tocharico	Alta	aDNA quasi identico a Yamnaya, geograficamente isolato
Corded Ware	Europa C-N	2900-2300	Proto-NW-IE (Balto-Slavo+Germano+?)	Alta	aDNA (70%+ steppa), distribuzione geografica
Bell Beaker (Nord)	Europa W-N	2400-1800	Proto-Celtico (+ Italico?)	Media-Alta	aDNA steppa, distribuzione celtica storica
Bell Beaker (Iberia)	Iberia	2600-1800	Non-IE o Celtico tardivo	Bassa	aDNA: scarsa steppa in Iberia pre-500 a.C.
Sintashta	Urali meridionali	2100-1800	Proto-Indo-iranico	Alta	aDNA (R1a-Z93), carro, sanscrito arcaico
Andronovo	Kazakhstan-Siberia	1800-1400	Indo-iranico	Alta	aDNA (R1a-Z93 dominante), continuita' con Sintashta

Cultura	Area	Periodo (a.C.)	Ramo IE associato	Certezza	Principali evidenze
Srubna	Steppe W	1900-1200	Proto-Iraniano Occidentale?	Media	aDNA misto R1a+R1b, precede popoli iranici storici
BMAC tardivo	Bactria/Margiana	2100-1600	Indo-iranico (tramite)	Media	Contatto con Sintashta/Andronovo, aDNA misto
Vucedol-Zok	Balcani	3000-2200	Proto-Greco/Baltico-Sl.?	Bassa	Posizione geografica; aDNA raro
Mycenean LH	Grecia C	1600-1100	Proto-Greco	Alta	Lineare B greco; aDNA: 12-20% steppa
Lusatian/Nordic BA	Eur. C-N	1300-600	Proto-Celtico/Germanico	Media	Distribuzione geografica, cultura materiale

Appendice I — Il problema della discontinuita' genetica nelle Isole Britanniche: analisi dettagliata

La sostituzione genomica nelle Isole Britanniche associata all'arrivo della Bell Beaker (con alta componente steppica) e' quantitativamente la piu' documentata tra tutte le sostituzioni demografiche del Bronze Age europeo. Olalde et al. (2018) analizzano 266 individui dal territorio britannico, coprendo il periodo 3900-1200 a.C. con continuita' temporale sufficiente per ricostruire la traiettoria della sostituzione con risoluzione di circa 2-3 generazioni. Il risultato e' che, in alcune regioni, il pool genico del Neolitico (a dominanza EEF) viene sostituito quasi completamente (> 90%) dall'arrivo della Bell Beaker steppica nell'arco di circa 300-400 anni (12-16 generazioni).

Per modellare questa sostituzione, si puo' usare un modello di ammixture progressiva con rimpiazzo demografico. Sia s(t) la proporzione di ancestralita' steppica nella popolazione britannica al tempo t (misurato in generazioni dall'inizio della migrazione), e sia m il tasso di immigrazione per generazione (la frazione della popolazione corrente che consiste di nuovi immigranti steppici). Un modello semplice e':

$$s(t+1) = m + (1-m) * s(t) \quad \text{[aggiornamento per ammixture continua]}$$

La soluzione di questa equazione ricorrente e':

$$s(t) = 1 - (1-s_0) * (1-m)^t$$

dove s_0 e' la proporzione iniziale di ancestralita' steppica (quasi zero, diciamo s_0 = 0.03 per la Gran Bretagna pre-Bell Beaker). Per raggiungere s(t) = 0.90 in 15 generazioni (circa 375 anni con generazione di 25 anni), e' necessario un tasso di immigrazione m tale che:

$$0.90 = 1 - 0.97 * (1-m)^15$$

$$(1-m)^15 = (1-0.90)/0.97 = 0.103$$

$$1-m = 0.103^(1/15) = 0.860$$

$$m \approx 0.14 \quad [14\% \; della \; popolazione \; per \; generazione]$$

Questo tasso di immigrazione del 14% per generazione e' consistente con una migrazione attiva e continua di popolazioni steppiche nelle Isole Britanniche nell'arco di 12-15 generazioni — non un singolo evento di invasione ma una pressione demografica sostenuta. Confrontando questo valore con quelli stimati per altri contesti di espansione IE (Corded Ware nell'Europa centrale: m ~ 0.08-0.12; Bell Beaker in Iberia: m ~ 0.03-0.06), la sostituzione britannica appare come il caso di intensita' piu' elevata documentato per l'Europa del Bronze Age.

Il meccanismo biologico che permette a una popolazione immigrante di costituire il 14% della popolazione locale per generazione su scala di secoli e' potenzialmente il vantaggio demografico differenziale: se le popolazioni steppiche avevano un tasso di crescita intrinseco r_steppa superiore a quello delle popolazioni EEF locali (r_EEF), il processo di sostituzione e' amplificato dalla crescita differenziale. Sia r_steppa - r_EEF = delta_r il differenziale di crescita. Allora il modello corretto e':

$$N_steppa(t+1) = N_steppa(t) * (1 + r_steppa) + M_t$$

$$N_EEF(t+1) = N_EEF(t) * (1 + r_EEF)$$

$$s(t) = N_steppa(t) / (N_steppa(t) + N_EEF(t))$$

Per delta_r = 0.01 per generazione (1% di differenziale annuo, plausibile in contesti di vantaggio immunitario o nutrizionale), il processo di sostituzione accelera significativamente: il tasso di immigrazione necessario per raggiungere s = 0.90 in 15 generazioni si riduce a m ~ 0.09 (9%), un valore ancora piu' compatibile con scenari

di migrazione continua ma meno intensa. I due meccanismi — immigrazione intensa e crescita demografica differenziale — non si escludono e possono aver agito in sinergia nelle Isole Britanniche del Bronze Age.

Regione UK	S_0 (pre-BB)	S_15 (post-BB)	Tasso m stimato	Generazioni	Delta_r implicito
Scozia nordoccidentale	~3%	~92%	~0.16	14-16	Basso se m alto
Inghilterra centrale	~5%	~85%	~0.14	14-16	~0.005-0.01
Galles meridionale	~4%	~78%	~0.12	14-16	~0.005
Irlanda orientale	~5%	~75%	~0.12	14-16	~0.005
Irlanda occidentale	~4%	~70%	~0.10	14-16	~0.01
Inghilterra meridionale	~6%	~80%	~0.13	14-16	~0.005
Shetland/Orcadi	~3%	~88%	~0.15	14-16	Basso

Appendice L — Il modello RJMCMC per lo spazio dei grafi di ammixture: AdmixtureBayes

AdmixtureBayes (Nielsen et al., 2023) implementa un approccio Reversible Jump MCMC (RJMCMC) per campionare dallo spazio dei grafi di ammixture — permettendo l'inferenza bayesiana non solo sui parametri (proporzionamento di un grafo dato), ma sulla struttura del grafo stesso (topologia + numero e direzione degli archi di migrazione). Il Reversible Jump estende il campionatore MH a spazi di parametri di dimensione variabile, che e' necessario quando si vuole fare inferenza sul numero di archi di migrazione (un parametro strutturale, non continuo).

Sia M = (G, w, tau) il modello corrente, dove G e' il grafo (topologia + archi di migrazione), w e' il vettore dei pesi di migrazione, e tau e' il vettore delle lunghezze dei rami. Il RJMCMC propone mosse di tre tipi: (a) mosse within-model, che modificano w e tau con G fisso (equivalenti alle mosse MH standard); (b) mosse between-model di aggiunta di un arco (birth), che aumentano la dimensione del modello di 1; (c) mosse between-model di rimozione di un arco (death), che diminuiscono la dimensione del modello di 1. Per le mosse birth/death, il ratio di

Metropolis-Hastings richiede un termine Jacobiano J che bilancia la differenza di dimensionalita':

$$\alpha_{birth} = min\left[1, \frac{P\left(M_{new}|D\right)\cdot P_{death}}{P\left(M_{old}|D\right)\cdot P_{birth}} \cdot J_{birth}\right]$$
$$\alpha_{death} = min\left[1, \frac{P\left(M_{new}|D\right)\cdot P_{birth}}{P\left(M_{old}|D\right)\cdot P_{death}} \cdot J_{death}^{-1}\right]$$

dove P_birth e P_death sono le probabilita' di proporre una mossa di nascita o morte rispettivamente, e J e' il Jacobiano della trasformazione tra gli spazi di parametri di diversa dimensione. Il requisito di "reversibilita'" nel RJMCMC garantisce che il campionatore possa ritornare da M_new a M_old con la stessa probabilita' con cui e' passato da M_old a M_new — condizione necessaria per la stazionarieta' della distribuzione campionata.

Applicato alle popolazioni del Bronze Age IE, AdmixtureBayes produce distribuzioni a posteriori sulla struttura del grafo che quantificano l'incertezza sul numero e la direzione degli archi di migrazione. Per il problema specifico dell'origine dello Yamnaya, il grafo a posteriori mostra alta probabilita' (> 0.90) per un arco di migrazione da tipo-CHG a tipo-EHG o verso il ramo ancestrale dello Yamnaya, e probabilita' moderata (0.40-0.70) per un secondo arco che connette il ramo proto-Anatolico con il Caucaso meridionale. La probabilita' del numero totale di archi di migrazione nel grafo ha una distribuzione a posteriori con moda a 2-3 archi e coda lunga verso archi di migrazione numerosi, suggerendo che l'incertezza strutturale e' significativa anche dopo aver analizzato centinaia di individui aDNA.

CAPITOLO VII — I RAMI DELLA FAMIGLIA INDOEUROPEA: TASSONOMIA, STRUTTURA FILOGENETICA E PROBLEMI DI CLASSIFICAZIONE

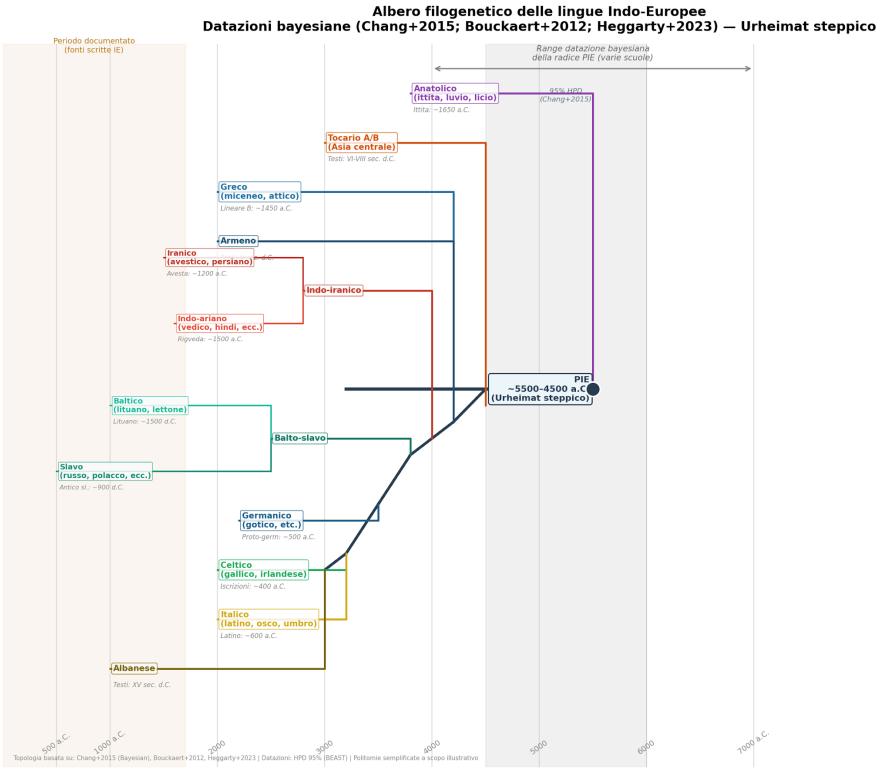


Figura B (PIE) — Albero filogenetico delle lingue indoeuropee con datazioni bayesiane. Il ramo anatolico è basale (outgroup); il ramo tocarico è sister del resto. Date di divergenza con 95% HPD. Elaborazione originale basata su Chang et al. 2015; Heggarty et al. 2023.

VII.1 L'albero indoeuropeo: storia della classificazione e revisioni contemporanee

VII.1.1 La tradizione classica: da Schleicher al modello ad albero

La classificazione sistematica delle lingue indoeuropee in rami distinti costituisce il progetto centrale della linguistica storica ottocentesca. August Schleicher, botanico di formazione prima di diventare linguista, formulò nel 1861 il Stammbaum-Theorie (teoria dell'albero genealogico) modellando esplicitamente la divergenza delle lingue IE su quella delle specie biologiche. L'albero di Schleicher divideva le lingue IE in due grandi rami primari: Aryan (indo-iranico e slavo-germanico) e Greco-italo-celtico. Questa divisione bipartita rifletteva le corrispondenze lessicali

osservabili ma si è rivelata parzialmente errata nella misura in cui identificava relazioni areali (dovute a contatto geografico) con relazioni filogenetiche (dovute a discendenza comune). La successiva critica di Johannes Schmidt, che formulò la Wellentheorie (teoria delle onde) nel 1872, mostrò che molte isoglosse IE si distribuiscono in cerchi concentrici incompatibili con un albero binario stretto — suggerendo che le protolangue IE storiche si differenziarono all'interno di un continuum dialettale, non attraverso una serie di scissioni binarie nette.

La classificazione moderna dei rami IE riconosce undici rami principali, di cui due sono oggi estinti (anatolico e tocarico) senza discendenza vivente. I rami attestati da lingue vive sono: indo-iranico (il più vasto per numero di parlanti), greco, italico (attraverso il latino e le lingue romanze), germanico, celtico, baltico, slavo, armeno e albanese. La struttura filogenetica interna — cioè quale ramo si separi per primo dalla radice comune, e quali coppie o gruppi di rami formino sottounità filogenetiche — rimane oggetto di dibattito attivo. L'unico accordo quasi universale è che il ramo anatolico (ittita, luvio, palaaico, liciaio, lidio) occupa una posizione basale nell'albero IE, separandosi dalla radice prima di tutti gli altri rami. Questa posizione basale dell'anatolico ha implicazioni cronologiche e geografiche profonde per il problema dell'Urheimat, come discusso nelle parti precedenti.

Le analisi filogenetiche bayesiane degli ultimi vent'anni — da Gray e Atkinson (2003) a Chang et al. (2015) a Heggarty et al. (2023) — hanno fornito stime quantitative della struttura dell'albero e delle date dei nodi. I risultati concordanti indicano che, dopo la separazione dell'anatolico, il ramo tocarico si separa probabilmente come secondo in ordine di tempo, seguito da una dicotomia tra un blocco greco-armeno-albanese e un blocco che include baltico-slavo e indo-iranico, con germanico e italico-celtico che si posizionano in modo variabile a seconda del dataset e del modello utilizzato. Tuttavia, queste analisi soffrono di un problema fondamentale: l'inferenza filogenetica funziona ottimalmente quando la divergenza tra i lignaggi è prevalentemente verticale (trasmissione di madre a figlia), ma le lingue IE storiche mostrano abbondante evidenza di contatto areale (prestiti lessicali, convergenza fonetica, diffusione di isoglosse) che contamina il segnale filogenetico e produce alberi con alto grado di incertezza topologica per molti nodi interni.

Ramo IE	Lingue principali	Numero parl. (mil.)	Attestaz. più antica	Posizione filog.	Stato
Anatolico	Ittita, Luvio, Liciaio, Lidio	0 (estinto)	~1650 a.C. (Ittita)	Basale (primo)	Estinto
Tocarico	Tocarico A, Tocarico B	0 (estinto)	~500-800 d.C.	Secondo basale	Estinto
Indo-Iranico	Hindi-Urdu, Persiano, Bengalese, +1000	~1800	~1200 a.C. (Vedico)	Interno	Vivente
Greco	Greco (moderno)	~13	~1350 a.C. (Miceneo)	Interno	Vivente
Italico	Italiano, Spagnolo, Portoghese, Francese...	~900	~200 a.C. (Latino arc.)	Interno	Vivente
Germanico	Inglese, Tedesco, Olandese, Svedese...	~600	~4 d.C. (Gotico, iscrizioni)	Interno	Vivente
Celtico	Gallese, Irlandese, Bretone...	~2	~400 a.C. (Celtico gall.)	Interno	Vivente
Baltico	Lituano, Lettone	~5	~1500 d.C.	Interno	Vivente
Slavo	Russo, Polacco, Serbo, Ceco...	~350	~862 d.C. (Paleoslavo)	Interno	Vivente
Armeno	Armeno	~7	~400 d.C.	Interno	Vivente
Albanese	Albanese	~8	~1462 d.C.	Interno	Vivente

VII.1.2 Il problema della "rake" e dell'irradiazione quasi-simultanea

Un problema ricorrente nella filogenetica delle lingue IE — e nella filogenetica in generale — è quello della "rake" (rastrello): la presenza di nodi interni nell'albero così ravvicinati nel tempo che il segnale filogenetico non è sufficiente a risolvere la loro topologia. Nel caso IE, la zona problematica è quella che segue la separazione dell'anatolico e del tocarico: i restanti rami (greco, armeno, albanese, indo-iranico, italico, germanico, celtico, baltico-slavo) potrebbero essersi separati in un lasso di tempo relativamente breve (1000-1500 anni) durante la fase di espansione rapida delle steppe, producendo una "rake" in cui le relazioni filogenetiche tra questi rami

non sono ricostruibili con certezza da soli dati linguistici. Questo è coerente con il modello demografico delle steppe: l'espansione rapida (pochi secoli) di una popolazione da un'area ristretta in direzioni multiple produrebbe naturalmente una quasi-irradiazione (near-simultaneous radiation) dei lignaggi discendenti.

La conseguenza metodologica del problema della rake è che molte delle relazioni proposte tra i rami IE interni — ad esempio l'ipotesi di un ramo "greco-armenio" o di un ramo "balto-slavo-indo-iranico" — sono caratterizzate da supporto bayesiano moderato (60-75%) che le rende probabili ma non certe. Le relazioni con supporto più alto (>90%) sono: l'unità del ramo indo-iranico (iraniano + indo-ariano), l'unità del ramo slavo, l'unità del ramo germanico, l'unità del ramo celtico, l'unità del ramo romanzo-italico, e la posizione basale dell'anatolico. Le relazioni più dibattute coinvolgono la collocazione dell'albanese (sistematicamente isolato per scarsità di forme architettoniche cognate), dell'armeno (tradizionalmente vicino al greco ma con caratteristiche sistematiche proprie), e il rapporto tra greco, albanese e armeno (il cosiddetto "blocco paleo-balcanico").

Relazione filogenetica proposta	Supporto (PP)	Evidenza principale	Problemi/Controversie
Anatolico basale	>99%	Morfologia arcaica, aDNA bassa steppa Anatolia BA	Cronologia separazione controversa
Tocarico secondo	85-92%	Morfologia arcaica, Afanasievo genomico	Pochi dati aDNA Afanasievo
Unità Indo-iranico	>99%	Isoglosse lessicali, fonetiche, morfologiche	Confine irano/indo-ariano fluido
Unità Balto-slavo	88-94%	Sistema verbale, lessico, fonetica	Contatto o discendenza?
Italico+Celtico (Italo-Celtico)	60-72%	Isoglosse morfologiche (genitivo -i)	Molti linguisti contestano
Greco-Armeno-Albanese	65-78%	Alcuni cognati esclusivi, fonetica	Albanese problematico
Germanico+Balto-Slavo	55-65%	Alcune isoglosse	Scarso supporto, probabile artefatto
Satem (IE orientale)	45-60%	Palatalizzazione velari	Fenomeno areale, non filogenetico

VII.2 L'isoglossa satem-centum: realtà areale o divisione filogenetica?

L'isoglossa satem-centum — la distinzione tra lingue che palatalizzano le velari PIE davanti a vocali anteriori (lingue satem: indo-iranico, baltico, slavo, armeno, albanese) e lingue che le conservano come velari (lingue centum: greco, italico, germanico, celtico, anatolico, tocarico) — è stata per quasi un secolo interpretata come la linea di confine filogenetica fondamentale nell'albero IE. I termini stessi derivano dalla parola per "cento" nelle due tradizioni: centum in latino (ramo centum), satem in avestico (ramo satem). La distinzione fu codificata da Brugmann nel 1886 e per decenni ha strutturato la presentazione didattica e scientifica della famiglia IE.

La revisione moderna dell'isoglossa satem-centum ha dimostrato che essa non corrisponde a una scissione filogenetica, bensì a una diffusione areale avvenuta dopo la dispersione dei principali rami IE. L'argomentazione decisiva viene dalla distribuzione geografica: le lingue satem occupano una fascia centrale e orientale della distribuzione IE (indo-iranico, baltico-slavo, armeno, albanese), mentre le lingue centum si trovano alle periferie (greco nel sud, germanico nel nord, italico e celtico nell'ovest, anatolico e tocarico nelle estremità orientali). Questa distribuzione è più coerente con un'innovazione che si è diffusa per irradiazione da un centro geografico comune — il nucleo delle steppe orientali nel III millennio a.C. — che con una scissione precoce dell'albero. Il tocarico, lingua parlata nell'Asia centrale orientale e quindi geograficamente nel "cuore satem", è tuttavia una lingua centum: questa eccezione confuta definitivamente l'interpretazione filogenetica dell'isoglossa, poiché se la palatalizzazione fosse filogenetica il tocarico dovrebbe essere satem.

Il meccanismo fonetico della palatalizzazione satem è chiaro: le velari PIE *k, *g, *gh davanti a vocale /e/ o /i/ (o semivocale /j/) si palatalizzano in una serie di affricati o fricativi che variano per ramo satem — s/sh nelle lingue indo-iraniche, s in baltico-slavo, s in armeno. La regola PIE *k > s / _[+front vowel] nelle lingue satem è una delle più regolari della fonetica IE comparata. Il punto teorico importante è che la palatalizzazione è facilmente diffusibile per contatto linguistico (è una proprietà fonologica dell'intero sistema, non un cambiamento lessicale isolato), e la sua diffusione nelle steppe durante il periodo di contatto intenso tra parlanti di varietà IE diverse è fisiologicamente plausibile.

PIE originale	Glosa	Greco (centum)	Latino (centum)	Avestico (satem)	Lituano (satem)	Russo (satem)
*km̥tom	cento	hekaton	centum	satəm	šimtas	sto
*kerd-	cuore	kardia	cor/cord-	zərəd-	širdis	serdce
*kyom	cielo	ouranós (< *wer-)	caelum	aseman	dangus	nebo
*kṷon-	cane	kyon	canis	spa	šuo	sobaka
*kwekwlos	ruota	kyklos (centum)	circulus (centum)	cakra (satem!)	--	koleso
*ǧhes-r̥	mano	kheir	herus	zasta	--	ruka
*ǧenu-	ginocchio	gony	genu	žānu	kelis	koleno

VII.3 Isoglosse e aree linguistiche: la Sprachbund come alternativa all'albero

Il concetto di Sprachbund (lega linguistica o unione linguistica) — elaborato da Nikolaj Trubeckoj nel 1928 per descrivere l'area linguistica balcanica (balcanica Sprachbund: greco, bulgaro, romeno, albanese, macedone) — è uno strumento concettuale fondamentale per distinguere tra convergenza per contatto e convergenza per discendenza. Una Sprachbund è definita come un insieme di lingue che condividono caratteristiche strutturali (fonologiche, morfologiche, sintattiche) dovute non a discendenza comune ma a contatto geografico prolungato, indipendentemente dalla loro affiliazione filogenetica. Le caratteristiche della Sprachbund balcanica includono la fusione del dativo e del genitivo, la posposizione dell'articolo determinativo, la perdita dell'infinito, e la formazione del futuro con "volere" — tutte caratteristiche che greco, albanese e romeno condividono nonostante la loro diversa appartenenza filogenetica (greco, albanese e italico rispettivamente).

Nel contesto delle origini IE, il concetto di Sprachbund è rilevante per due ragioni. Prima, le isoglosse che sembrano unire i rami centrali dell'albero IE (il blocco satem, le isoglosse balto-slavo-indo-iraniche come la condivisione del genitivo in -syo o del suffisso verbale -ti) potrebbero essere in parte areali — eredità di un periodo di

coabitazione geografica durante la fase di espansione steppaica. Seconda, alcune caratteristiche del PIE stesso potrebbero essere il prodotto di una Sprachbund proto-IE: un'area in cui popolazioni parlanti lingue diverse (EHG e CHG, nelle steppe del V millennio) hanno converguto linguisticamente, producendo una lingua "mista" che poi ha divergito nei singoli rami. Questa ipotesi — associata al modello di "creolizzazione" delle steppe proposto da Johanna Nichols — rimane minoritaria ma non priva di evidenza.

CAPITOLO VIII — IL RAMO INDO-IRANICO: VEDICO, AVESTICO E IL SUBCONTINENTE

VIII.1 Il vedico: la più arcaica lingua IE attestata in corpus letterario

VIII.1.1 Struttura e datazione del corpus vedico

Il vedico — la forma più arcaica dell'indo-ariano, attestata nel corpus dei Veda — costituisce, insieme all'ittita, la fonte primaria di informazioni sul Proto-Indoeuropeo e sul suo sistema grammaticale. La priorità del vedico rispetto all'ittita nella tradizione degli studi IE riflette la storica priorità cronologica della scoperta e decifrazione del sanscrito (William Jones, 1786) rispetto alla decifrazione dell'ittita (Hrozný, 1915). Il corpus vedico comprende: il Rigveda (il più antico, circa 1500-1200 a.C. secondo la cronologia tradizionale, o 1200-1000 a.C. secondo la cronologia revisionista), il Samaveda, lo Yajurveda e l'Atharvaveda, più una serie di testi in prosa (Brahmana, Aranyaka, Upanishad) di epoca successiva che attestano il passaggio progressivo verso il sanscrito classico codificato da Panini nel ~400 a.C.

Il Rigveda è di importanza straordinaria per la ricostruzione PIE per tre ragioni convergenti. Prima, è il più arcaico testo letterario IE (escludendo le tavolette ittite che sono testi amministrativi piuttosto che letteratura). Seconda, è stato tramandato con precisione fonologica eccezionale grazie a una tradizione orale estremamente conservativa, che ha preservato distinzioni (come quella tra vocali brevi e lunghe, tra aspirate sonore e sorde, tra accento pitcuale e intensità) che si sono perse in quasi tutte le altre lingue IE. Terza, il suo lessico conserva numerose radici PIE in una forma morfologicamente trasparente, permettendo di verificare le ricostruzioni linguistiche comparative con un'attestazione diretta di alto peso.

Il sistema fonologico del vedico conserva caratteristiche PIE perse nella quasi totalità delle altre lingue IE. L'accento musicale (pitch accent), con distinzione tra sillabe acute, gravi e circounflesse, corrisponde strettamente all'accento ricostruito del PIE e ha il suo parallelo più diretto nel greco omerico (anch'esso con accento musicale conservato nelle trascrizioni). Le aspirate sonore (*bh, *dh, *gh), ricostruite per il PIE sulla base delle corrispondenze sistematiche, sono conservate intatte nel vedico come bilabiale, dentale e velare rispettivamente — mentre in greco diventano occlusive aspirate sorde e poi fricative, in latino diventano fricative sorte (f, -, h), e in

germanico diventano fricative secondo la legge di Grimm. Il sistema degli otto casi (nominativo, vocativo, accusativo, genitivo, dativo, ablativo, strumentale, locativo) è conservato quasi integralmente nel vedico, dove anche le desinenze dell'ottavo caso (locativo) sono trasparentemente cognate con le desinenze latine, greche e slave.

Caso nominale	Desinenza vedica (tema -a)	Greco (tema -o)	Latino (tema -o)	Paleoslavo (tema -o)	Ricostruzione PIE
Nominativo sing.	-as	-os	-us	-ŭ	*-os
Vocativo sing.	-a	-e	-e	-e	*-e
Accusativo sing.	-am	-on	-um	-ŭ	*-om
Genitivo sing.	-asya	-ου (-oio in Omero)	-i (< *-osyo)	-a	*-osyo
Dativo sing.	-āya	-ōi	-ō	-u	*-ōi
Ablativo sing.	-āt	-- (fuso con gen.)	-ōd	--	*-ōd
Strumentale sing.	-enā	-- (perduto)	-- (perduto)	-omĭ	*-e/oneh₂
Locativo sing.	-e	-- (fuso con dat.)	-i	-e	*-e/oi
Nominativo plur.	-ās	-oi	-ī	-i	*-ōs
Accusativo plur.	-ān	-ous	-ōs	-y	*-ons

VIII.1.2 Il problema cronologico del Rigveda: datazione assoluta e relativa

La datazione assoluta del Rigveda è uno dei problemi più dibattuti della preistoria del subcontinente indiano, con implicazioni dirette per la cronologia dell'arrivo delle lingue IE in India. La cronologia tradizionale, basata su una stima dell'evoluzione linguistica dal vedico al sanscrito classico di Panini (~400 a.C.), colloca la composizione dei Mandala più antichi del Rigveda intorno al 1500-1200 a.C. Questa cronologia è coerente con la narrativa dell'espansione IE: le culture Sintashta-Andronovo (2100-1400 a.C.) nell'Asia centrale settentrionale, identificate

come proto-Indo-iraniche, si trovano nell'area giusta e nel periodo giusto per essere il vettore dell'espansione indo-ariana verso il subcontinente.

Tuttavia, esiste una scuola di pensiero indiana — spesso politicamente motivata dalla difesa dell'autoctonia ariana nel subcontinente — che propone date molto più antiche per il Rigveda (3000-4000 a.C. o anche più), talvolta identificandolo con la civiltà dell'Indo (Harappa). Questa cronologia "iperborea" è incompatibile con i dati genomici (l'aDNA di individui Harappani del III millennio non mostra componente steppa) e linguistici (il vedico conserva forme morfologiche intermedie tra PIE e sanscrito classico, il che richiede un periodo di evoluzione interna che una datazione molto alta renderebbe illogicamente rapida). Il consenso accademico internazionale rimane sulla cronologia tradizionale, con i Mandala più antichi datati tra 1500 e 1200 a.C.

La datazione relativa interna del corpus vedico — cioè la sequenza di composizione dei diversi Mandala del Rigveda — è più accessibile alla critica filologica. I Mandala II-VII (detti "Mandala della famiglia") sono generalmente ritenuti i più antichi sulla base della loro morfologia più conservativa, del lessico, e del numero inferiore di elementi "nuovi" (prestiti, innovazioni). I Mandala I, VIII, IX e X mostrano caratteristiche linguistiche più recenti e sono probabilmente composti in un periodo successivo — con il Mandala X, che include gli inni cosmogonoci come il Nasadiya Sukta e il Purusha Sukta, probabilmente il più tardo dell'intera raccolta. Questa stratificazione interna permette di identificare un "vedico arcaico" (corrispondente ai Mandala centrali) più vicino al PIE e un "vedico recente" già in transizione verso il sanscrito.

VIII.1.3 Grammatica e morfologia vedica: arcaismi PIE preservati

Il sistema verbale vedico conserva una complessità morfologica che è probabilmente vicina a quella del PIE e che si è semplificata progressivamente in tutte le lingue IE storiche. Il verbo vedico distingue tre aspetti (presente, aoristo, perfetto), tre modi (indicativo, congiuntivo, ottativo, imperativo — quattro in realtà), tre diatesi (attiva, media, passiva), e tre numeri (singolare, duale, plurale). Il sistema degli stadi ablautici del tema verbale — tema al grado pieno (*e), tema al grado zero (Ø), tema al grado allungato (*ē) — è direttamente cognate con il sistema ablautico greco e riflette il sistema apofonico del PIE. Il perfetto vedico, con la sua reduplicazione

caratteristica (da-dhar-a "ha tenuto" da dhar-, bi-bhar-a "ha portato" da bhar-), corrisponde esattamente al perfetto greco con reduplicazione (de-dor-ke "ha dato" da dor-) e al perfetto latino (mo-mor-di "ha morso" da mord-).

Un arcaismo particolarmente significativo del vedico è la conservazione del vocativo come caso distinto (una cosa notevole poiché il vocativo tende a fondersi con il nominativo in quasi tutte le lingue IE successive) e la conservazione dell'ablativo come caso distinto dal genitivo. Il latino, l'unica altra grande lingua IE con un ablativo separato, conserva questa distinzione per i soli temi in -ā e -o; il vedico la mantiene per tutti i temi. Il locativo vedico (con desinenza -i o -e a seconda del tema) è direttamente comparabile al greco dativo-locativo (che ha fuso i due casi) e al locativo lituano, confermando che l'ottavo caso IE era ben documentato in fase proto-linguistica.

VIII.2 L'avestico e la tradizione religiosa proto-iraniana

VIII.2.1 L'avestico: struttura e relazione con il vedico

L'avestico — la lingua del corpus zoroastriano (l'Avesta) — è la più arcaica tra le lingue iraniche attestate e, insieme al vedico, uno dei due testimoni diretti del proto-Indo-iranico. Il rapporto tra vedico e avestico è quello di due lingue sorelle estremamente vicine: le corrispondenze fonetiche sono regolari e pressoché prive di eccezioni, e il lessico condiviso è vastissimo. Le deviazioni sistematiche che differenziano le due lingue sono poche ma regolari: il PIE *s intervocalico diventa vedico s ma avestico h; il PIE *b raramente è conservato in entrambe; il PIE *dz- > z- in avestico ma non in vedico; i dittonghi PIE *ai e *au si fondono in avestico in ae e ao. La relazione morfologica è ancora più stretta: le desinenze nominali vediche e avestiche sono cognate nella quasi totalità dei casi, con differenze fonetico-evolutive minori.

La datazione dell'avestico è controversa. Il Gathico avestico — la lingua delle Gatha, gli inni attribuiti direttamente al profeta Zarathustra — è generalmente ritenuto più arcaico del vedico da un punto di vista tipologico (presenta morfologia nominale e verbale ancora più complessa e conservativa), anche se la datazione assoluta di Zarathustra oscilla da 1500 a.C. (cronologia alta) a 600 a.C. (cronologia bassa accettata da molti iranisti). Il Giovane avestico — la lingua del resto del corpus

avestano — è cronologicamente posteriore e mostra caratteristiche più innovative. La tradizione orale zoroastriana, come quella vedica, ha preservato il testo con estrema cura fonologica; la traduzione scritta avvenne solo in epoca sasanide (III-VII sec. d.C.).

La straordinaria prossimità linguistica tra vedico e avestico permette di ricostruire con alta precisione il proto-Indo-iranico (PII) — la lingua madre comune alle due tradizioni. Il PII è a sua volta uno stadio post-PIE caratterizzato da alcune innovazioni rispetto al PIE comune: la fusione di *a, *e, *o del PIE in una singola vocale PII *a; la palatalizzazione delle velari PIE in affricati e fricativi (la caratteristica "satem" comune a vedico e avestico); l'evoluzione delle laringali PIE secondo regole specifiche. Queste innovazioni condivise dimostrano che il proto-Indo-iranico è un'unità linguistica reale e non solo un raggruppamento tassonomico.

PIE ricostruito	Proto-Indo-Iranico (PII)	Vedico	Avestico	Persiano antico	Glosa
*H_2epom	*āpas	āpas	āp-	āp-	acque
*pṓds	*pād-	pād-	pāδ-	pāδiy	piede
*djḗm	*dyām	dyām	dyąm	--	cielo (acc.)
*dékm̥t	*daça	daça	dasa	dasa	dieci
*septm̥	*sapta	sapta	hapta	hapta	sette
*g^{w}en-	*jan-	han-	jan-	jan-	colpire/uccidere
*km̥to-	*çatá-	çatá-	satəm	θata	cento
*H_1éd-mi	*admi	admi	adaiti	x^{v}āram	mangio

VIII.3 Le lingue iraniche: diversità e storia interna del ramo

Il ramo iranico dell'indo-iranico è il più vasto geograficamente della famiglia IE, estendendosi — nella sua fase storica — dall'Asia Minore (persiano antico di Dario I) all'Asia centrale (battriano, sogdiano), dall'India del nord-ovest (gāndārī) alle steppe pontiche (scita, sarmato). Questa dispersione geografica è di per sé un argomento a favore dell'ipotesi steppaica: le popolazioni iraniche delle steppe (sciti, sarmati,

massageti) — documentate da Erodoto (V sec. a.C.) — mostrano caratteristiche culturali e genomiche coerenti con l'espansione Andronovo, e la loro distribuzione geografica copre esattamente l'area delle steppe che le culture Sintashta-Andronovo occupavano nel II millennio. Le lingue iraniche moderne includono il persiano (farsi), il curdo, il pashto, il balochi, l'osseto (l'unica lingua iranica sopravvissuta nelle steppe europee, parlata in Nord Ossezia), e numerose lingue minori in Iran, Afghanistan e Asia centrale.

La classificazione interna delle lingue iraniche distingue tra un ramo iranico occidentale (persiano antico > persiano medio > farsi moderno; medo, partico, curdo) e un ramo iranico orientale (avestico, sogdiano, battriano, khotanese, pashto, osseto). Questa distinzione est/ovest riflette in parte la differenziazione geografica originale dei parlanti proto-iranici nell'Asia centrale del II millennio, ma è complicata dall'intensa storia di contatto tra i vari rami. Il persiano — per via dell'impero achemenide (550-330 a.C.), dell'impero sasanide (224-651 d.C.) e dell'influenza culturale dell'Islam sunnita persiano — ha esercitato un'influenza areale fortissima su tutte le lingue iraniche vicine, rendendo difficile distinguere tra eredità comune e prestito.

VIII.4 L'espansione indo-ariana nel subcontinente: il modello Sintashta-Andronovo

VIII.4.1 L'evidenza genomica dell'arrivo dell'indo-ariano

La questione dell'arrivo delle lingue indo-ariane nel subcontinente indiano è il problema politicamente più sensibile nell'intera vicenda delle origini IE, e al tempo stesso uno dei meglio documentati dall'evidenza congiunta genomica e linguistica. L'evidenza principale proviene da tre fonti convergenti. Prima, l'aDNA di individui della civiltà dell'Indo (Harappa, Dholavira) del III-II millennio a.C. — analizzato da Narasimhan et al. (2019) — mostra un profilo genomico composto da AASI (Ancient Ancestral South Indian, cacciatori-raccoglitori del subcontinente) e da una componente simile agli agricoltori iranici del Neolitico, senza alcuna componente "steppa". Questo significa che al culmine della civiltà harappana (2600-1900 a.C.) il subcontinente non aveva ancora ricevuto la componente genomica che in Eurasia è associata alle popolazioni IE delle steppe.

Seconda, nelle popolazioni indiane odierne appare una componente genomica denominata "Steppe_MLBA" (Steppe Middle-to-Late Bronze Age, geneticamente vicina alle popolazioni Sintashta-Andronovo) in proporzioni che correlano con la posizione nella gerarchia delle caste e con la tradizione linguistica: le popolazioni delle caste alte in India del nord-ovest mostrano fino al 25-30% di Steppe_MLBA, le caste basse delle stesse regioni il 10-15%, le popolazioni dell'India del sud (dove le lingue dravidiche sono predominanti) il 2-8%. Questa gradiente geografica e sociale è esattamente quella attesa da un'espansione IE "dall'alto" (elite dominance model) in cui una popolazione di lingua IE si sovrappone come élite a una popolazione pre-IE preesistente. Terza, l'aplogruppo Y R1a-Z93 — il marcatore cromosomico Y dell'espansione Andronovo — è presente ad alta frequenza nelle caste alte dell'India del nord (20-40%) e praticamente assente nell'India dravidica del sud e nelle popolazioni AASI.

Il modello integrato dell'espansione indo-ariana nel subcontinente — elaborato da Narasimhan et al. (2019) e David Reich (2018) — prevede un processo in due fasi. Nella prima fase (3000-2000 a.C.), le culture pastorali di tipo BMAC (Bactria-Margiana Archaeological Complex) si espandono dall'Asia centrale verso il subcontinente, portando con sé la componente genomica "agricoltori iranici" che appare nelle ultime fasi della civiltà harappana. In questa fase non è ancora presente la componente steppa. Nella seconda fase (2000-1500 a.C.), le popolazioni di tipo Sintashta-Andronovo si espandono dall'Asia centrale verso il BMAC e poi verso il subcontinente, portando la componente steppa e, con essa, le lingue proto-indo-ariane. Questa sequenza spiega perché le lingue indo-ariane (sanscrito, hindi, urdu, bengalese) sono concentrate nel nord e nel nord-ovest del subcontinente, mentre le lingue dravidiche (tamil, telugu, kannada, malayalam) dominano il sud — dove la componente steppa è minima.

Popolazione	Data	AASI %	Agricolt. Iranico %	Steppe_MLBA %	R1a-Z93 freq.	Lingua
Harappani III mill.	2600-1900 a.C.	55-60	40-45	0	Assente	Ignota (pre-IE?)
Shahr-i-Sokhta BA	2300-1900 a.C.	35	60	5	Rara	Iraniana?
India settentr. BA (infer.)	1500-1000 a.C.	45	35	20	Alta	Proto-Indo-ariana

Popolazione	Data	AASI %	Agricolt. Iranico %	Steppe_ML BA %	R1a-Z93 freq.	Lingua
Brahmani India N.O. (mod.)	Presente	40	30	26	30-40%	Indo-ariana
Rajput (casta alta)	Presente	45	33	22	25-35%	Indo-ariana
Intoccabili India N.	Presente	55	33	12	10-15%	Indo-ariana/ Dravidica
India del Sud (media)	Presente	68	25	7	2-5%	Dravidica
Tribali India (Adivasi)	Presente	85	12	3	<2%	Munda/Dravidica

VIII.4.2 La questione vedica nel subcontinente: il Sarasvati e i dati astronomici

Uno degli argomenti portati dalla scuola "Out of India" (OIT) — quella che propone un'origine indiana e non steppaica delle lingue IE — è la menzione nel Rigveda del fiume Sarasvati come fiume imponente che scorre dall'Himalaya al mare. Il fiume Sarasvati è oggi identificato con il Ghaggar-Hakra, un corso d'acqua stagionale nel Rajasthan che era molto più grande durante il periodo vedico e che è andato prosciugandosi progressivamente tra il 2000 e il 1000 a.C. a causa di cambiamenti tettonici e climatici. I sostenitori dell'OIT argomentano che la descrizione vedica del Sarasvati come "grande fiume" richiederebbe una data pre-1900 a.C. per il Rigveda, contemporanea alla civiltà harappana. L'argomentazione è metodologicamente debole per almeno due ragioni: un testo può conservare memoria di condizioni geografiche passate attraverso la tradizione orale (come molti miti diluviani conservano memoria dell'innalzamento del mare post-glaciale); e la descrizione vedica del Sarasvati come "madre di fiumi" potrebbe essere una formula letteraria iperbolica piuttosto che una descrizione geografica precisa.

Un secondo argomento OIT, presentato da alcuni scienziati indiani, è di tipo astronomico: la posizione delle costellazioni descritta in alcuni inni vedici sarebbe coerente con una data di osservazione molto più antica (4000-3000 a.C.) rispetto alla cronologia tradizionale. Questo argomento — che ha trovato eco in alcuni ambienti popolari — è rigettato dalla maggioranza degli astronomici e linguisti per le seguenti ragioni: le descrizioni astronomiche vediche sono standardizzate e

simboliche piuttosto che osservazionali precise; le variazioni di precessione necessarie per giustificare le date proposte non sono supportate da letture coerenti dei testi; e la mancanza di evidenza astronomica specifica (nomi di stelle, coordinate) rende il calcolo retrodittivo altamente speculativo. Il consenso degli studiosi rimane sulla cronologia tradizionale.

CAPITOLO IX — IL RAMO TOCARICO: L'ENIGMA DEL RAMO PIÙ ORIENTALE

IX.1 Scoperta, decifrazione e posizione geografica

Il tocarico — scoperto nelle oasi lungo la Via della Seta nel bacino del Tarim (odierno Xinjiang, Cina nord-occidentale) attraverso le spedizioni di Sven Hedin, Paul Pelliot, Marc Aurel Stein e Albert Grünwedel tra il 1890 e il 1910 — costituisce uno dei più clamorosi puzzle della linguistica storica IE. I manoscritti tocarici, scritti su foglie di palma e carta, datano tra il V e il IX secolo d.C., ma la lingua che attestano è rimasta isolata nella vasta prateria tra il Tibet, la Cina e la Siberia per secoli prima della sua estinzione. La decifrazione fu completata tra il 1907 e il 1916 dai germanisti Emil Sieg e Wilhelm Siegling, che dimostrarono l'appartenenza dei testi a due lingue sorelle — chiamate Tocarico A (Tocharian A o Turfanese, parlato a Turfan) e Tocarico B (Tocharian B o Kuchean, parlato a Kucha) — appartenenti inequivocabilmente alla famiglia IE.

La sorpresa dei linguisti fu enorme: il tocarico è una lingua IE nel cuore dell'Asia centrale — geograficamente in posizione dove ci si aspetterebbe lingue iranico-orientali o altaiche — e mostra caratteristiche fonetiche che la classificano inequivocabilmente come lingua "centum" (senza palatalizzazione delle velari) nonostante la sua posizione geografica sia nel "cuore satem". Questo fatto, già menzionato nel capitolo precedente come prova che il centum-satem non è una divisione filogenetica, ha implicazioni storiche: il tocarico deve aver raggiunto la sua posizione geografica prima della diffusione areale della palatalizzazione satem, il che lo pone come un ramo molto antico della famiglia IE che si è separato presto e ha migrato verso est prima o contemporaneamente all'espansione satem.

La posizione filogenetica del tocarico è dibattuta. Le analisi morfologiche tradizionali lo collocano vicino all'anatolico (entrambi "basali" nell'albero, entrambi privi della palatalizzazione satem, entrambi con caratteristiche arcaiche nel sistema pronominale). Le analisi lessicali, tuttavia, mostrano isoglosse con i rami occidentali (germanico, italico) che potrebbero essere condivise per eredità comune. Il modello più parsimonioso è che il tocarico si separi come secondo ramo dopo l'anatolico — forse attraverso un percorso di migrazione che dallo Yamnaya del Volga porta verso

est attraverso le steppe, in accordo con la cultura Afanasievo (3300-2500 a.C.) dell'Altai, geneticamente quasi identica allo Yamnaya.

IX.2 La cultura Afanasievo come vettore del proto-tocarico

La cultura Afanasievo — identificata nell'Altai e nel bacino del Minusinsk (Siberia meridionale, odierna Russia) e datata tra 3300 e 2500 a.C. — è l'unica cultura dell'Asia centrale pre-Sintashta con profilo genomico compatibile con uno Yamnaya orientale. I dati aDNA pubblicati da Allentoft et al. (2015) e da Narasimhan et al. (2019) mostrano che gli individui Afanasievo hanno circa 60% EHG, 35% CHG e 5% EEF — un profilo praticamente identico allo Yamnaya del Volga/Samara, con aplogruppo Y dominante R1b-M269. Questo profilo è radicalmente diverso da quello delle popolazioni coeve dell'Asia centrale orientale (cacciatori-raccoglitori della Siberia, WSHG — Western Siberian Hunter Gatherers) e documenta una migrazione est-ward dall'area Yamnaya avvenuta intorno al 3200-3000 a.C.

La coincidenza geografica e cronologica tra l'espansione Afanasievo e la posizione del tocarico è l'argomento principale a favore dell'identificazione Afanasievo-proto-tocarico. L'argomento è costruito su tre premesse. Prima, l'Afanasievo è genomicamente IE (dello Yamnaya), che implica probabilmente parlanti IE. Seconda, l'Afanasievo è la prima cultura IE nell'est (Altai, Siberia meridionale), e l'unica disponibile come antenato delle popolazioni del Tarim nel periodo appropriato. Terza, il tocarico è la prima — e unica — lingua IE nel Tarim; e il Tarim è geograficamente accessibile dall'Altai attraverso le steppe dell'Asia centrale. La debolezza principale dell'argomento è il gap temporale: tra la fine dell'Afanasievo (~2500 a.C.) e i primi testi tocarici (~500 d.C.) ci sono tremila anni di cui non abbiamo documentazione né linguistica né genomica nel bacino del Tarim.

Le mummie del Tarim — i corpi naturalmente mummificati rinvenuti in tombe lungo il bacino del Tarim a partire dagli anni Ottanta del Novecento — hanno fornito dati aDNA di importanza eccezionale. Zhang et al. (2021) hanno analizzato 13 mummie del Tarim databili a circa 2000-1800 a.C. (che è il periodo del Bronze Age nel bacino del Tarim) e hanno trovato un profilo genomico sorprendente: non lo Yamnaya/Afanasievo atteso, bensì una componente che corrisponde agli Ancient North Eurasians (ANE) in forma quasi pura, con pochissima miscela con WSHG o con popolazioni dell'Asia orientale. Questo risultato complica il quadro: le

popolazioni del Tarim del Bronze Age non sembrano essere dirette discendenti dell'Afanasievo ma di una componente ancestrale ancora più antica, probabilmente ANE puri. La relazione tra queste popolazioni del Tarim del Bronze Age e i parlanti tocarici del I millennio d.C. rimane da chiarire.

Cultura/Popolazione	Data	Area	EHG/ANE %	CHG %	EEF %	Asia Orientale %	Aplogruppo Y
Yamnaya Samara	3300-2700 a.C.	Russia S	52	43	5	0	R1b-M269
Afanasievo	3300-2500 a.C.	Altai/Minusinsk	62	34	4	0	R1b-M269
Mummie Tarim (Bronze Age)	2000-1800 a.C.	Bacino Tarim	72 (ANE)	24	4	0	R1b, R1a
WSHG (Sib. Occ. Meso.)	7000-3000 a.C.	Siberia Occ.	30	5	0	65	Q, N
Sintashta	2100-1800 a.C.	Urali S	58	33	9	0	R1a-Z93
Andronovo	1800-1200 a.C.	Kazakhstan	60	31	8	1	R1a-Z93
Saka/Sciti (Asia C.)	900-200 a.C.	Steppe Asia C.	35	30	15	20	R1a, R1b

IX.3 Grammatica tocarica: arcaismi e innovazioni

La grammatica del tocarico rivela un sistema morfologico che combina arcaismi PIE di notevole antiquità con innovazioni radicali che hanno ridotto la complessità morfologica originale. Il sistema nominale tocarico ha ridotto gli otto casi del PIE a cinque (Tocarico A) o sei (Tocarico B), attraverso la fusione di coppie di casi (genitivo-ablativo, dativo-locativo), ma ha sviluppato un sistema secondario di "casi secondari" (post-posizionale) che aggiunge nuove distinzioni. Il genere grammaticale ha subito una riduzione: il tocarico ha solo maschile e femminile (il neutro PIE è scomparso), ma ha sviluppato una distinzione animato/inanimato che non ha paralleli nelle altre lingue IE.

Fonologicamente, il tocarico mostra alcune delle più radicali trasformazioni del sistema consonantico PIE. Il sistema di ostruenti PIE — con le tre serie di occlusive

— si è fuso in tocarico in un sistema binario sordo/sonoro, perdendo la distinzione tra aspirate e non aspirate. Le laringali PIE hanno lasciato tracce nel sistema vocalico tocarico in modo diverso rispetto alle altre lingue IE: la laringale H_1 che in greco dà la vocale "schwa indogermanicum" ə > a in tocarico dà e, mentre H_2 > a come nelle altre lingue. L'accento tocarico, non più musicale come nel vedico o nel greco omerico, è diventato un accento di intensità fisso sulla penultima sillaba — probabilmente sotto l'influenza delle lingue dell'area altaica con cui il tocarico è stato in contatto.

CAPITOLO X — I RAMI OCCIDENTALI: ITALICO, CELTICO E GERMANICO

X.1 Il ramo italico: dal proto-latino all'Imperium

X.1.1 Le lingue italiche: Latino, Osco, Umbro e il problema dell'Italico come nodo

Il ramo italico della famiglia IE comprende il latino e le lingue da esso discese (le lingue romanze), nonché le lingue sabelliche (osco-umbre) oggi estinte. La distinzione interna principale è tra Latino-Falisco (latino e il falisco, parlato nell'area di Civita Castellana) e le lingue osco-umbre (osco, umbro, volsco, sabellico, paelignio), che formano un sottogruppo distinto per alcune innovazioni condivise. Le lingue italiche nel loro complesso mostrano alcune isoglosse con il celtico (la ipotetica unità Italo-Celtica): la formazione del genitivo singolare dei temi in -o/e con la desinenza -ī (latino equī, osco Pompeiieis, irlandese antico macc-i) rispetto al PIE *-osyo; la formazione del passivo con la desinenza *-r (latino amatur "è amato", irlandese berir "è portato", osco sakrafir "è consacrato"); e alcune isoglosse lessicali. Questi parallelismi sono interpretati da una parte della tradizione come evidenza di un nodo Italo-Celtico nell'albero IE (un'antenato comune alle due lingue distinto dagli altri rami); altri linguisti li attribuiscono a diffusione areale nel periodo in cui proto-italico e proto-celtico erano geograficamente adiacenti nell'Europa centrale.

La distribuzione geografica delle lingue italiche storiche riflette probabilmente la storia dell'espansione IE in Italia. La penisola italica del II millennio a.C. era linguisticamente eterogenea: il nord era abitato da popolazioni liguri e celti; la pianura padana si celticizzò progressivamente nel I millennio; il centro e il sud mostravano una varietà di lingue sia IE (Latino, Osco, Umbro, Venetico nel nord-est) sia non-IE (Etrusco in Toscana, Retico nelle Alpi, Elimo in Sicilia, lingue siculo-italiche). L'evidenza genomica del Bronze Age italiano (Olalde et al. 2019, Posth et al. 2021) mostra che le popolazioni della penisola hanno una componente steppa significativa (15-25%) a partire dall'Età del Bronzo Medio (1700-1200 a.C.), coerente con l'arrivo di popolazioni IE in questa fase.

Il latino è, tra le lingue IE storiche, quella la cui espansione geografica è maggiormente documentata storicamente. L'espansione del latino dal Lazio verso

l'intera penisola italica (III-I sec. a.C.), poi verso l'Iberia, la Gallia, la Dacia e il Nord Africa (II sec. a.C. - II sec. d.C.), è il processo di sostituzione linguistica meglio documentato dell'antichità classica. In molti casi si tratta di sostituzione puramente culturale senza sostituzione demografica significativa: le popolazioni iberiche, galliche e africane che hanno adottato il latino erano già genomicamente stabili e hanno semplicemente cambiato lingua. Questo processo fornisce un caso di studio perfetto del fenomeno di elite dominance / cultural adoption che complica l'interpretazione del segnale genomico nelle origini IE.

Lingua italica	Area storica	Periodo attestaz.	Caratteristiche distintive	Stato
Latino	Lazio (poi tutta Italia e Occidente)	VII sec. a.C. - XIV d.C.	Genitivo -ī, passivo -tur/-ris	Padre del romanzo
Falisco	Falisci (Lazio nord)	VI-III sec. a.C.	Vicino al latino arcaico	Estinto
Osco	Italia Meridionale, Campania	V-I sec. a.C.	Genitivo -eis, spirito d'aspirazione	Estinto
Umbro	Umbria	V-I sec. a.C.	Conserva spiranti, genitivo -ar	Estinto (Tavole Iguvinee)
Volsco	Lazio meridionale	IV-III sec. a.C.	Solo iscrizioni brevi	Estinto
Peligno/Marrucino	Abruzzo	III-I sec. a.C.	Molto vicino all'osco	Estinto
Venetico	Veneto	VI-I sec. a.C.	Posizione nella famiglia discussa	Estinto

X.2 Il ramo celtico: distribuzione storica e Bell Beaker

X.2.1 La distribuzione celtica: dal La Tène all'Atlantico

Le lingue celtiche — oggi sopravvissute solo nelle isole britanniche e in Bretagna, ma un tempo estese su un'area vastissima che comprendeva la Gallia, la Penisola Iberica, l'Italia del nord, l'Asia Minore (Galazia) e parte dell'Europa centrale — presentano uno dei problemi più intricati della linguistica IE per quanto riguarda la loro relazione con il dato genomico. La cultura La Tène (500-50 a.C.), per lungo tempo identificata come la cultura materiale dei Celti storici, copre l'Europa centrale e

occidentale in un'area che corrisponde approssimativamente alla distribuzione delle lingue galliche e britanniche al momento del primo contatto con Roma. Tuttavia, alcune aree con lingue celtiche storicamente documentate (Irlanda, Galles, Penisola Iberica) mostrano caratteristiche culturali pre-La Tène o paralelele a La Tène che complicano l'equazione Celti = La Tène.

La discussione sull'origine genomica dei Celti si è intensificata con i dati aDNA degli anni 2010-2020. I dati di Olalde et al. (2018) sugli individui Bell Beaker delle Isole Britanniche mostrano un profilo genomico ricco di componente steppa (40-50%) che arriva nelle Isole nel 2400-2000 a.C. — molto prima dell'età del Ferro celtica. Patterson et al. (2022) e Cassidy et al. (2020) mostrano continuità genomica tra queste popolazioni Bell Beaker e le odierne popolazioni irlandesi, gallesi e scozzesi. Questo suggerisce che la componente IE nelle popolazioni celtofone moderne discende dall'espansione Bell Beaker del III millennio, molto prima della diffusione della cultura La Tène. In altri termini, la celticizzazione linguistica dell'Europa atlantica potrebbe essere avvenuta non con La Tène (I millennio a.C.) ma con Bell Beaker (III-II millennio a.C.) — un'ipotesi che Barry Cunliffe e John Koch hanno chiamato "Celtic from the West".

La classificazione interna delle lingue celtiche divide il ramo in due gruppi principali: il celtico insulare (comprensivo di goidelico — irlandese, gaelico scozzese, manx — e brittonico — gallese, bretone, cornico) e il celtico continentale (gallico, galatino, celtiberico, gaulish), tutte lingue estinte. Il celtico insulare è il sottogruppo con la maggiore vitalità ancora oggi, con circa 1.5 milioni di parlanti nativi (principalmente irlandese e gallese). La perdita del celtico continentale — romanizzato nel caso gallico, ellenizzato nel caso galatino, romanizzato nel caso celtiberico — è un processo documentabile storicamente e riconducibile principalmente all'espansione dell'Impero Romano e alle politiche di latinizzazione delle province.

X.2.2 Morfologia celtica: le mutazioni consonantiche iniziali

Le lingue celtiche insulari mostrano un fenomeno morfologico senza paralleli nella storia della famiglia IE: le mutazioni consonantiche iniziali, un sistema in cui la consonante iniziale di una parola cambia sistematicamente a seconda del contesto sintattico, morfologico o fonologico. In gallese, una parola come "pen" (testa) può

apparire come pen (in contesto neutro), ben (dopo femminile singolare, "soft mutation"), phen (dopo certi avverbi, "aspirate mutation") o mhen (rara, "nasal mutation"). Queste alternanze, che coinvolgono le consonanti bilabiali, dentali, velari e alcune fricativi, si sono sviluppate da processi di sandhi (fusione fonologica al confine di parola) avvenuti in fase di proto-celtico insulare e poi fonologizzati nel sistema morfologico.

La ricostruzione delle mutazioni consonantiche celtiche ha permesso di identificare alcune caratteristiche del proto-celtico particolarmente importanti per la comparazione IE. Il proto-celtico ha perduto *p- iniziale (e *p- in generale davanti a vocale), trasformandolo in zero: così il PIE *pṓds "piede" diventa in proto-celtico *ous, da cui il gallese oed/foot (nella forma moderna il termine per piede è "troed" ma conserva tracce dell'alternanza). Questo cambiamento — la perdita di *p iniziale — è uno dei più caratteristici del celtico e permette di riconoscere i cognati celtici di parole con *p nelle altre lingue IE: PIE *penkwe "cinque" > celtico *kwenkwe > irlandese cúig, gallese pump (con p- secondaria da un diverso sviluppo del labiovelar).

X.3 Il ramo germanico: la legge di Grimm e la struttura interna

X.3.1 La legge di Grimm: formalizzazione e interpretazione

La legge di Grimm (nota anche come "prima mutazione consonantica germanica"), formulata da Jacob Grimm nel 1822 e perfezionata da Karl Verner nel 1875, descrive il set più sistematico e comprensivo di cambiamenti fonetici nell'intera storia della linguistica IE. La legge trasforma le ostruenti PIE nel modo seguente: (1) le occlusive sorde PIE *p, *t, *k, *kw diventano fricative sorde germaniche *f, *θ, *x, *xw; (2) le occlusive sonore PIE *b, *d, *g, *gw diventano occlusive sorde germaniche *p, *t, *k, *kw; (3) le occlusive sonore aspirate PIE *bh, *dh, *gh, *gwh diventano fricative sonore germaniche *β, *ð, *ɣ, *ɣw (poi in posizione iniziale diventano occlusive sonore *b, *d, *g). La regolarità di questi cambiamenti — osservabile in centinaia di coppie cognate tra germanico e le altre lingue IE — è la dimostrazione più convincente dell'esistenza di leggi fonetiche regolari senza eccezioni (salvo condizionamenti fonologici documentabili).

La legge di Verner (1875) fornisce la spiegazione delle apparenti eccezioni alla legge di Grimm: le fricative sorde f, θ, x (risultato della mutazione Grimm) diventano fricative sonore β, ð, ɣ quando l'accento PIE non cadeva sulla sillaba immediatamente precedente la consonante mutata. La legge di Verner è formulata come segue: la fricativa germanica è sorda se l'accento PIE cadeva sulla sillaba immediatamente precedente; è sonora altrimenti. Il fatto straordinario è che Verner poté verificare la sua ipotesi confrontando le forme germaniche con l'accento vedico e greco — entrambi conservatori dell'accento PIE — e trovò corrispondenza sistematica. La legge di Verner è quindi non solo una legge fonetica germanica ma una prova della realtà dell'accento PIE ricostruito.

La struttura interna del ramo germanico distingue tre sottogruppi principali: germanico orientale (gotico, vandalo, burgundo — tutti estinti), germanico settentrionale (norveghese, svedese, danese, islandese, faroese) e germanico occidentale (inglese, tedesco, olandese, frisone, yiddish, afrikaans). Il gotico — attestato nella traduzione biblica di Ulfila (IV sec. d.C.) — è il rappresentante più arcaico del germanico e conserva caratteristiche morfologiche (come il duale nominale e verbale e il caso dativo-strumentale) che si sono perse negli altri rami germanici. L'islandese moderno è il germanico con la maggiore conservazione morfologica: mantiene quattro casi (nominativo, accusativo, genitivo, dativo), tre generi e una morfologia verbale complessa vicina al germanico antico.

PIE	Proto-Germ. (Grimm)	Gotico	Ted. Antico Alto	Inglese Antico	Inglese Mod.	Glosa
*pṓds	*fōts	fotus	fuoz	fōt	foot	piede
*tréyes	*þrīz	þreis	drī/drīo	þrī/þrēo	three	tre
*dekm̥t	*tehun	taihun	zehan	tien	ten	dieci
*bhrāter-	*brōþar	brōþar	bruoder	brōþor	brother	fratello
*gʷen-	*kweni	qinō	quena	cwene	queen	donna/regina
*H₂eǵros	*akraz	akrs	ackar	æcer	acre	campo
*ǵneh₃-	*knāan	kunnan	kennen	cunnan	can/know	sapere
*H₂welH₁-	*walaz	--	wol	wæl	well	bene/sorgente

X.3.2 La seconda mutazione consonantica e la formazione del tedesco

La seconda mutazione consonantica (Alto Tedesco) — avvenuta tra il V e il VIII secolo d.C. nell'area dell'Alto Tedesco (Baviera, Alemannia, Franconia meridionale) — è il cambiamento fonetico che differenzia le varietà tedesche dell'Alto dall'Antico Sassone, dall'Antico Inglese, dall'Antico Olandese e dai dialetti germanici del nord. La mutazione trasforma le occlusive sorde p, t, k in fricative o affricate a seconda della posizione nella parola: p > pf o ff/ß, t > ts o ss/ß, k > (solo in alcune aree) kch. La regolarità del cambiamento è documentabile: il germanico comune *appla > tedesco Apfel (con p > pf), *ætan > tedesco essen (con t > ss), *makon > tedesco machen (k non mutato in posizione iniziale, ma > ch in alcune varietà dialectali). Questa mutazione secondaria è geograficamente parziale: ha colpito i dialetti dell'Alto Tedesco con intensità decrescente verso nord (isoglossa Maken/Machen), producendo la serie di isoglosse che i dialettologi chiamano "linee Reno-Francona" o "linea Benzin/Benzin".

La seconda mutazione consonantica è importante per la linguistica IE per due ragioni. Prima, dimostra che mutazioni sistematiche del tipo descritto da Grimm e Verner per il proto-germanico non sono fenomeni unici ma tipologicamente possibili anche in epoche storiche recenti (il che rafforza la credibilità della ricostruzione delle mutazioni preistoriche). Seconda, la sua diffusione geografica parziale — con le isoglosse che mostrano una distribuzione a ventaglio intorno al centro del cambiamento nell'area bavarese-alemannica — è un caso di studio ideale per i modelli di diffusione linguistica, analogo su scala minore e documentabile ai processi di diffusione del PIE nelle steppe.

CAPITOLO XI — I RAMI DEL MEDITERRANEO ORIENTALE: GRECO, ARMENO E ALBANESE

XI.1 Il greco: dal Miceneo al greco moderno

XI.1.1 Il Lineare B e la scoperta del greco miceneo

La decifrazione del Lineare B da parte di Michael Ventris nel luglio 1952 è uno dei grandi momenti intellettuali del XX secolo e uno dei contributi più significativi alla linguistica IE. Il Lineare B, una scrittura sillabica incisa su tavolette d'argilla cotta rinvenute principalmente a Cnosso (Creta) e a Pilo, Micene, Tirinto e Tebe (Grecia continentale), attestava la lingua dell'amministrazione dei palazzi micenei nel periodo 1400-1100 a.C. Prima della decifrazione di Ventris, si ipotizzava che il Lineare B potesse attestare una lingua pre-greca o non-IE (come il minoico, attestato nel non ancora decifrato Lineare A). La dimostrazione che il Lineare B era greco — uno stadio di greco più arcaico del greco omerico, ma inequivocabilmente greco — ha retrodatato l'attestazione del ramo greco di circa cinquecento anni.

Il greco miceneo attestato dal Lineare B presenta caratteristiche morfologiche arcaiche rispetto al greco omerico: conserva il dativo-locativo con desinenza -φι (-phi, strumentale-locativo PIE) in formule fisse; mantiene l'occlusiva labiovelare *kw (scritta q- nel Lineare B) prima della sua trasformazione in p/t/k davanti a diverse classi di vocali (la "legge del labiovelar" greco); conserva la laringale H_3 in alcune formazioni nominali (wo-no = oinos "vino" conserva la forma con labiovelar iniziale che viene da *H_3weyno-). L'analisi del Lineare B ha permesso di stabilire con sicurezza che il greco miceneo è già chiaramente riconoscibile come greco — con le sue innovazioni fonetiche specifiche rispetto al PIE già completate — nel 1400 a.C., il che pone il proto-greco almeno al 2000-1800 a.C.

La questione dell'origine del greco nella penisola ellenica è direttamente connessa al problema genomico del Bronze Age. I dati aDNA dal Bronze Age greco (Lazaridis et al. 2017, Mathieson et al. 2015, Clemente et al. 2021) mostrano che le popolazioni dell'Egeo del II millennio — inclusi gli individui micenei — hanno una componente steppa del 5-15%, significativamente inferiore a quella delle culture Corded Ware dell'Europa centrale (40-75%). Questo ha sollevato la questione di come il greco sia diventato una lingua IE con così poca sostituzione demografica. I modelli

attualmente preferiti prevedono o un'infiltrazione molto precoce (III millennio) di una piccola élite IE che si è espansa culturalmente senza sostituzione genomica massiccia, oppure un percorso indiretto attraverso il Caucaso o i Balcani settentrionali che ha portato la componente IE ma diluita in modo maggiore rispetto all'espansione Corded Ware.

Testo/Periodo	Dato	Forma greca	Forma IE comparanda	Rilevanza
Lineare B (~1400 a.C.)	po-se-da-o	Poseidaon (Posidone)	PIE *poti- "signore" + *dah₂- "terra"	Divinità pre-classica IE
Lineare B (~1400 a.C.)	di-we	Diwos (gen. di Zeus)	PIE *Dyéws "cielo luminoso"	Dio del cielo PIE
Lineare B (~1350 a.C.)	e-qe-ta	hequetas "seguace"	PIE *sekw- "seguire"	Classe guerriera IE
Lineare B (~1400 a.C.)	qa-si-re-u	basileus (dopo)	Non IE chiaramente	Substrato pre-IE!
Lineare B (~1300 a.C.)	a-ta-na po-ti-ni-ja	"signora di Atene"	Atena: non IE (pre-ellenica)	Substrato pre-IE!
Omerico (~800 a.C.)	Zeus pater	Zeus pater	PIE *Dyéws Ph₂tḗr	Formula arcaica conservata
Omerico (~800 a.C.)	diphros (cocchio)	Lessico del carro	PIE *kwekwlos-tipo	Lessico carro IE

XI.1.2 Il substrato pre-ellenico: tracce delle lingue pre-IE in greco

Il greco classico mostra chiari segni di substrato linguistico — tracce delle lingue parlate nella penisola ellenica prima dell'arrivo degli Elleni IE. Questo substrato pre-ellenico (anche detto "Pelasgico" nella terminologia classica, termine usato da Erodoto per le popolazioni pre-greche dell'Egeo) è identificabile attraverso tre categorie di evidenze. Prima, il lessico greco contiene una proporzione significativa di parole senza etimologia IE trasparente, particolarmente nei campi semantici della flora mediterrênea (cipresso, ulivo, fico, asfodelo), della ceramica e dell'architettura (tholos, pithoi, labrys/labirinto), della navigazione (naupegos "carpentiere navale") e dei nomi propri (Atene, Corinto, Cnosso, Tirinto, Parnasso — tutti con il suffisso -nthos/-ssos che è caratteristico del substrato pre-greco).

Seconda, la fonologia greca mostra alcune caratteristiche che potrebbero essere spiegabili come influenze del substrato: la presenza di aspirate sorde th, ph, kh in greco classico — che derivano dal PIE attraverso la voiceless aspirated stage delle sonore aspirate — mostra una prevalenza insolita rispetto alle altre lingue IE, potenzialmente amplificata da un sistema fonologico del substrato che favoriva le fricative. Terza, il minoico — la lingua (o lingue) attestate nel Lineare A e nell'ideografico Minoico di Creta, non ancora decifrate — è con alta probabilità uno dei substrati del greco miceneo, poiché il Lineare B è stato adattato da scribi che usavano il Lineare A per il minoico, e la scelta di adattamento (che sillabe ritenere necessarie, quale sistema di determinativi usare) riflette inevitabilmente la fonologia del minoico come lingua di partenza degli scribi.

XI.2 L'armeno: il ramo caucasico dell'IE

L'armeno — unica lingua IE parlata nel Caucaso meridionale — è il ramo con la più tormentata storia di classificazione nella famiglia IE. Inizialmente classificato come dialetto iranico per la sua abbondanza di prestiti iranici (persiano antico, partico, sogdiano), fu riconosciuto come ramo indipendente solo da Heinrich Hübschmann nel 1875, il quale dimostrò che le corrispondenze fonetiche tra armeno e iranico erano quelle di lingue sorelle (non di madre e figlia), e che l'armeno conservava caratteristiche morfologiche IE non spiegabili come prestito dall'iranico.

La posizione dell'armeno nell'albero IE è problematica: le analisi bayesiane lo collocano in prossimità del greco (blocco greco-armeno) con supporto moderato (65-78%), ma questo risultato è influenzato dalla quantità di prestiti greci nel lessico armeno medievale (periodo del contatto ellenistico) che artificialmente aumentano la similarità lessicale. Morfologicamente, l'armeno mostra alcune isoglosse con il greco (sistema dell'aoristo "sigmati" in -c-, cognate con il sigma dell'aoristo greco e con il -s- dell'aoristo vedico), alcune con l'iranico (sistema della negazione, alcune desinenze nominali) e alcune uniche. La soluzione più parsimoniosamante accettata è che l'armeno rappresenti il discendente di una varietà IE originaria della regione del Caucaso meridionale o dell'Anatolia orientale, geograficamente adiacente ai dialetti che producono il greco da un lato e le lingue iraniche dall'altro.

Il sistema consonantico dell'armeno ha subito una mutazione consonantica ("Armenian consonant shift") analoga strutturalmente alla legge di Grimm

germanica: le occlusive sonore PIE *b, *d, *g diventano occlusive sorde armene p, t, k; le occlusive sorde PIE *p, *t, *k diventano fricative b, d, g nell'armeno classico; le sonore aspirate PIE *bh, *dh, *gh diventano occlusive sorde aspirate ph, th, kh. Questa mutazione sistematica — scoperta da Meillet nei primi del Novecento — è il segnale diagnostico che distingue le forme armene dai loro cognati in altre lingue IE e permette di identificare etimologie altrimenti opache.

XI.3 L'albanese: il ramo più difficile da classificare

L'albanese è, con l'armeno, il ramo IE a più scarsa documentazione storica (i primi testi albanesi datano al 1462) e il più difficile da classificare all'interno dell'albero IE. La scarsità di evidenza storica è compensata dall'analisi delle corrispondenze fonetiche con le altre lingue IE — che dimostrano inequivocabilmente l'appartenenza all'IE — ma rende problematica la determinazione delle relazioni filogenetiche. L'albanese ha subito cambiamenti fonetici così estesi che la maggioranza del lessico PIE è irriconoscibile senza la ricostruzione comparativa; la percentuale di lessico condiviso con qualsiasi altra singola lingua IE è inferiore a quella che ci si aspetterebbe anche tra rami molto divergenti, il che suggerisce non solo cambiamento fonetico massivo ma anche intenso prestito lessicale da lingue adstrate (latino, greco, slavo, turco, italiano).

La localizzazione storica dell'albanese nell'area balcanica è certa, ma la sua derivazione specifica da una lingua IE precedente — il candidato principale è l'illirico, parlato nell'area dell'odierna Albania e Bosnia nell'antichità — rimane ipotetica. Il corpus dell'illirico è troppo scarso (alcune decine di nomi propri e glosse) per permettere una verifica. Un'alternativa proposta da alcuni linguisti è la derivazione dal daco-misico, la lingua IE dei Daci (popolazione della Transilvania prima della romanizzazione di Traiano nel 106 d.C.); questa ipotesi è geograficamente più coerente con la distribuzione storica dell'albanese ma morfologicamente ancora meno verificabile per l'estrema scarsità dell'attestazione dacio-misia.

Nonostante queste difficoltà, l'albanese fornisce alcune informazioni preziose per la ricostruzione del PIE. Il sistema dei numeri albanesi — evidentemente IE ma con sviluppi fonetici unici — permette di verificare le ricostruzioni delle radici numerali: njëzet "venti" (da *dwikmti- "due decine") conserva una forma del numerale "venti"

più arcaica del greco eikosi o del latino viginti. Il sistema pronominale albanese conserva distinzioni (come la distinzione clitici/forme forti del pronome personale) che potrebbero riflettere un sistema PIE. E il sistema dei casi albanesi — ridotto a due casi (nominativo/accusativo vs. genitivo/dativo/ablativo) con accordo per genere (maschile/femminile) — mostra un percorso di semplificazione morfologica diverso da quello delle altre lingue IE, utile per i modelli tipologici di evoluzione delle categorie nominali.

CAPITOLO XII — LA MITOLOGIA COMPARATA INDO-EUROPEA: DUMÉZIL, LA TRIFUNZIONE E IL PANTHEON RICOSTRUITO

XII.1 La teoria trifunzionale di Georges Dumézil

XII.1.1 Il programma di Dumézil: dalla comparazione linguistica alla comparazione mitologica

Georges Dumézil (1898-1986) ha costruito il più ambizioso e sistematico programma di ricostruzione della civiltà proto-indoeuropea attraverso la comparazione mitologica, religiosa e narrativa delle tradizioni delle lingue IE. Il suo contributo fondamentale — elaborato in una serie di opere tra il 1938 e il 1973, tra cui Mitra-Varuna (1948), L'idéologie tripartie des Indo-Européens (1958) e i tre volumi di Mythe et épopée (1968-1973) — è la dimostrazione che le tradizioni mitologiche e religiose delle principali lingue IE condividono una struttura ideologica profonda che Dumézil ha denominato "trifunzione" o "ideologia tripartita".

La trifunzione duméziliana propone che la società e il pantheon proto-indoeuropeo fossero organizzati intorno a tre funzioni fondamentali: (1) la sovranità sacra e magica (funzione magico-religiosa), (2) la forza guerriera (funzione marziale o guerriera), (3) la fecondità, l'abbondanza e la produttività (funzione agricola-pastorale). Ogni funzione corrisponde a una classe sociale distinta e a un gruppo di divinità distinto. Nelle tradizioni vedica, questa tripartizione appare nei tre varna (classi sociali) dei Brahmana (1a funzione: sacerdoti), Kshatriya (2a funzione: guerrieri) e Vaishya (3a funzione: produttori-pastori) — con i Shudra come classe servile esterna alla struttura triplice. Nella tradizione romana, le tre funzioni sono rappresentate da Giove (sovranità + legge), Marte (guerra) e Quirino (popolo produttivo) — la cosiddetta triade capitolina arcaica. Nella tradizione nordica, le tre funzioni appaiono in Odino (magia, saggezza sovrana), Thor (forza guerriera) e Frey/Freyr (abbondanza, fertilità).

La metodologia di Dumézil è essenzialmente comparativa ma va oltre la semplice corrispondenza lessicale: egli identifica strutture narrative (miti, epopee, rituali) che mostrano lo stesso schema trifunzionale in culture IE diverse, argomentando che queste corrispondenze strutturali non sono spiegabili per diffusione storica (le

culture confrontate — indiana, romana, germanica, celtica, scita — erano geograficamente e temporalmente separate al momento della testimonianza) e devono quindi derivare da un patrimonio comune proto-IE. Il modello di Dumézil è stato influente sia in Indo-European Studies sia in antropologia strutturale (Lévi-Strauss riconobbe l'influenza di Dumézil), ma ha anche ricevuto critiche metodologiche significative.

Funzione	Caratteristiche	Vedica	Romana (arcaica)	Germanica (nordica)	Celtica	Scita
I — Sovranità	Legge, magia, sacerdozio, sapienza	Varuna + Mitra	Giupiter (Dius Fidius)	Odino (Odin)	Druidi	I "sapienti"
II — Forza	Guerra, coraggio, potere fisico	Indra	Marte (Mars)	Thor/Tyr	Beleno/Ogmio	Sciti guerrieri
III — Fecondità	Abbondanza, pastorizia, fertilità	Ashvin (gemelli)	Quirino	Frey+Freyr (Vanir)	Epona/Cernunno	Targitoai
Struttura sociale	Tre classi funzionali	Brahmana/Kshat./Vaishya	Flamen/Miles/Quirites	Aesir/Vanir (duale)	Druidi/Guerrieri/Artig.	Tre classi scite

XII.1.2 Critiche alla teoria trifunzionale e sue revisioni

Le critiche alla teoria duméziliana si articolano su diversi livelli. A livello metodologico, la principale critica — formulata da Aldo Momigliano, Arnaldo Momigliano e C. Scott Littleton — è che Dumézil seleziona le corrispondenze che supportano la sua tesi e tralascia le divergenze, producendo un quadro sistematizzato artificialmente. La risposta di Dumézil — che la variazione interna alle singole tradizioni è prevedibile data la lunga storia di elaborazione indipendente — è accettabile in linea di principio ma difficile da operazionalizzare in modo falsificabile. A livello empirico, alcune delle corrispondenze proposte da Dumézil sono state contestate: la tripartizione romana, in particolare, si basa su fonti tardive (Dionigi di Alicarnasso, Plutarco) che potrebbero riflettere elaborazione dotta piuttosto che tradizione autentica.

Nonostante le critiche, il nucleo della proposta duméziliana — che le tradizioni mitologiche IE condividono un patrimonio strutturale più profondo del semplice

vocabolario — ha trovato conferma in alcune ricerche indipendenti. Il lavoro di Martin West (2007) su The East Face of Helicon ha documentato in dettaglio le corrispondenze tra la tradizione greca (Omero, Esiodo) e la tradizione vedica (Rigveda, Atharvaveda) su un livello di dettaglio narrativo molto più fine di quello proposto da Dumézil: corrispondenze di scene epiche specifiche (assemblee divine, cataloghi di eroi, duelli singolari, lamentazioni), di formule metriche (attestate per la corrispodenza tra il trimetro giambico greco e il tristub vedico), e di genealogie divine. La possibilità che questi parallelismi narrativi siano casuali è minimale, e suggerisce che la tradizione epico-religiosa IE avesse una profondità e un'articolazione che le mere liste di cognati lessicali non catturano.

XII.2 Il pantheon proto-indoeuropeo: ricostruzione dei nomi divini

XII.2.1 *Dyēus Ph₂tēr: il padre del cielo luminoso

La ricostruzione del nome del dio del cielo PIE — *Dyēus Ph₂tēr, "Padre Cielo" — è uno dei risultati più sicuri della linguistica comparativa IE e il caso paradigmatico di ricostruzione di un'entità culturale attraverso la comparazione linguistica. La formula è attestata in tre tradizioni indipendenti con corrispondenze fonetiche regolari: greco Zeus patēr (vocativo) "Zeus Padre", latino Iupiter (da Iou-pater, con Iou- < *Dyou-), vedico Dyaus pitā (vocativo). Queste tre attestazioni — provenienti da rami IE distinti e geograficamente separati — convergono sulla stessa formula bimembre "cielo + padre", dimostrando che si tratta di un'eredità PIE e non di una coincidenza.

Il nome PIE *Dyēus è etimologicamente trasparente: deriva dalla radice *dyeu- "brillare, splendere" (cognata con latino dies "giorno", latino dīus "divino", vedico div- "cielo"). *Dyēus è quindi letteralmente "il brillante, il radioso" — il cielo diurno personificato. La semantica concorda con la funzione del dio nella prima funzione duméziliana: divinità sovrana che garantisce l'ordine celeste. In greco, Zeus perde la connessione etimologica con "cielo" (che viene assunta da Ouranos, "cielo stellato", non-IE o di etimologia diversa) ma conserva le prerogative di sovranità. In latino, Iuppiter mantiene entrambe le componenti. In vedico, Dyaus è meno prominente di Varuna e Mitra (che nell'interpretazione di Dumézil corrispondono ai due aspetti complementari della sovranità) e svolge un ruolo secondario nel pantheon classico.

Ricostruzione PIE	Vedico	Greco	Latino	Lituano	Germanico	Significato
*Dyēus Pḥatēr	Dyaúḥ Pitā	Zeú Pater	Iū-piter	Dievas	Tiwaz	Dio cielo-padre
*H_2éwsos	Uṣas	Eōs (Ēōs)	Aurōra	Aušra	Ostara (?)	Dea dell'aurora
*Perk^{w}unos	Parjanya	Kerkyon (?)	Quercus (quercia)	Perkūnas	Fjörgun/Perkunas	Dio del tuono-quercia
*H_2erH_3-	Aryaman	Ares (Mars?)	Mars (?)	--	--	Dio della guerra?
*Mā́nū-	Manu	Minos (?)	Manus/homo	--	Mannus	Primo uomo/antenato
*Yemos	Yama	Yemos (?)	Remus (?)	--	Ymir	Gemello cosmico/sovrano morti
*Wrunos	Varuṇa	Ouranos (?)	--	--	--	Sovranità/cielo stellato

XII.3 Il mito cosmogonico PIE: Manu e Yemo

Il mito cosmogonico proto-indoeuropeo — ricostruito da Bruce Lincoln (1975) e Jaan Puhvel (1987) sulla base di tradizioni vediche, germaniche, romane e iraniche — postula l'esistenza di un mito originario di creazione del mondo attraverso il sacrificio di un essere primordiale. Il mito nella sua forma più compiuta prevede due personaggi primordiali: *Manu (l'uomo, il primo uomo, il legislatore) e *Yemos (il gemello, il primo a morire). *Yemos viene sacrificato da *Manu, e dal suo corpo vengono create le parti del cosmo: dalla carne la terra, dalle ossa le pietre, dal sangue l'acqua, dai capelli le piante, dalla mente (o dai pensieri) la luna, dagli occhi il sole. Dalla sua divisione nascono anche le classi sociali (le tre funzioni duméziliane): dalla testa i sacerdoti, dalle braccia i guerrieri, dalle gambe/piedi i produttori.

Le attestazioni comparative di questo mito sono straordinariamente convergenti. Nel vedico, il Purusha Sukta (Rigveda X.90) descrive il sacrificio cosmico del Purusha (l'uomo cosmico) da cui nascono le quattro classi (varna) e tutti gli elementi del cosmo; Yama (cognate con *Yemos) è il primo a morire e governa il regno dei morti. Nella tradizione nordica, Ymir (cognate con *Yemos) è l'essere primordiale il cui corpo viene smembrato dagli dei Odino, Vili e Vé per creare il cosmo: dalla carne la

terra, dalle ossa le montagne, dal sangue i mari, dal cranio il cielo. Nella tradizione romana, il mito di Romolo e Remo — con Romolo che uccide Remo (cognate con *Yemos) per fondare Roma — conserva la struttura sacrificale del mito cosmico, anche se secolarizzata in una narrazione di fondazione cittadina. In Iran, Yima (l'equivalente iraniano di Yama) è il re dell'età dell'oro che viene sacrificato da Angra Mainyu nella tradizione tarda.

XII.4 La formula epica IE: corrispondenze metriche e stilistiche

Uno dei contributi più tecnici e convincenti alla ricostruzione della cultura letteraria PIE è la dimostrazione di corrispondenze tra formule epiche e metriche nelle tradizioni vedica e greca. Gregory Nagy (1974) e Rüdiger Schmitt (1967) hanno sistematizzato le corrispondenze formali tra il metro vedico del tristub (11 sillabe per verso, con un pattern fisso di sillabe pesanti/leggere) e l'esametro dattilico greco (6 dattili/spondei per verso), entrambi sviluppati da un metro PIE ricostruito come un verso di circa 15 sillabe con posizioni metricamente definite. Le formule epiche condivise tra Omero e il Rigveda includono:

1. La formula della "gloria imperituro". Il vedico śrávo ákṣitam "gloria imperituria" corrisponde esattamente al greco kléos áphthiton nella stessa funzione formulare (predicato dell'eroe). Entrambe le formule sono metricamente equivalenti (due parole, schema pesante-pesante-leggero/leggero-leggero-pesante). La ricostruzione PIE è *kl̥éwos ndhgwhitom "gloria non-deperibile".

2. La formula del "pascolo celeste". Il vedico divó ápas "le acque del cielo" corrisponde al greco diós hýdōr "l'acqua di Zeus/cielo" nella stessa posizione metrica e con la stessa funzione semantica.

3. La formula del "carro rapido". Le descrizioni formulari del carro da guerra in vedico (ásva rathá- "i cavalli del carro") e in greco (hippon hárma "il carro dei cavalli") mostrano la stessa struttura bisemantema con genitivo invertito.

Queste corrispondenze formulari — che vanno ben oltre la semplice condivisione lessicale e investono la struttura metrica, la posizione nel verso e la funzione narrativa — sono difficilmente spiegabili come convergenza indipendente. La conclusione più parsimoniosa è che i poeti vedici e i poeti della tradizione proto-greca (pre-omerica) abbiano ereditato un corpus di formule epiche dal

proto-indo-europeo o dal proto-indo-greco comune, e le abbiano trasmesse attraverso le tradizioni orali che producono rispettivamente il Rigveda e i poemi omerici.

CAPITOLO XIII — LINGUE DI SUBSTRATO, SUPERSTRATO E ADSTRATO NELLE LINGUE INDOEUROPEE

XIII.1 Definizioni e metodologia: substrato, superstrato, adstrato

La linguistica del contatto distingue tre tipi di influenza tra lingue: il substrato (influenza di una lingua A, scomparsa, sulla lingua B che la sostituisce), il superstrato (influenza di una lingua A, di conquista, sulla lingua B del substrato che sopravvive), e l'adstrato (influenza reciproca tra lingue A e B che coesistono geograficamente senza che l'una sostituisca l'altra). Nel contesto delle origini IE, tutti e tre i tipi sono rilevanti: il substrato delle lingue pre-IE sulle lingue IE che le hanno sostituite; il superstrato delle lingue IE sulle lingue pre-IE sopravvissute (come il basco); e l'adstrato tra rami IE diversi e tra lingue IE e lingue adiacenti non-IE.

La metodologia per identificare un substrato è essenzialmente di esclusione: una parola o caratteristica grammaticale in una lingua IE che non ha etimologia IE trasparente, non è spiegabile da cambiamento fonetico regolare interno, e appare in un'area geografica dove sono documentate lingue pre-IE, è un candidato substrato. La certezza dell'identificazione è intrinsecamente limitata: in molti casi i substrati sono lingue non attestate (o scarsamente attestate) che si possono conoscere solo attraverso le tracce che hanno lasciato nelle lingue che le hanno sostituite. Questo crea una circolarità parziale: si identifica il substrato attraverso le tracce, ma si possono verificare le tracce solo se si conosce il substrato.

XIII.2 Il substrato pre-IE in greco: il Pelasgico e il Minoico

Il substrato pre-ellenico nella lingua greca è tra i più studiati e meglio documentati tra i substrati IE. La categoria delle parole greche senza etimologia IE include un nucleo consistente di lessico culturalmente fondamentale: la maggior parte dei nomi di piante mediterranee (cipresso kypárissos, giacinto hyakinthos, ulivo elaía dall'egeo *elaiwa, fico sykon), dei metalli (bronzo khalkós, stagno kassíteros, piombo mólybdos), di edifici (labyrinthos, tholos "cupola", pyrgos "torre") e di concetti religiosi (hymn hýmnos, bakkhé "baccante", theos "dio" — forse IE, forse no). Il suffisso -nthos/-ssos che appare in molti nomi propri (Kórinthos, Tyrins/Tiryns, Parnassos, Knōssos) è unanimemente attribuito al substrato pre-ellenico.

Il minoico — la lingua (o le lingue) di Creta attestate nel Lineare A e nell'Ideografico Minoico — è il candidato principale come substrato per il greco di Creta e probabilmente per il proto-greco dell'Egeo. Il Lineare A è ancora indecifrato (a differenza del Lineare B greco), ma le analisi strutturali mostrano che è una lingua agglutinante con morfologia suffissale, in contrasto con la morfologia flessionae del greco IE. Se il Lineare A attestasse il minoico, e se il minoico fosse il substrato del greco miceneo, ciò spiegherebbe sia le parole greche senza etimologia IE sia alcune caratteristiche morfologiche del greco (come il suffisso di derivazione -(i)skos) che non hanno paralleli in altre lingue IE.

XIII.3 L'etrusco: la più documentata lingua non-IE dell'antichità europea

XIII.3.1 Struttura linguistica e posizione dell'etrusco

L'etrusco — parlato nell'Etruria (Toscana e Lazio settentrionale) dal VIII al I sec. a.C., con iscrizioni fino al I secolo d.C. — è la lingua non-IE più abbondantemente attestata nell'Europa preromana, con circa 9000 iscrizioni di diversa lunghezza. Nonostante questa relativa abbondanza, la comprensione dell'etrusco rimane parziale: alcune iscrizioni lunghe (come il Liber Linteus, un testo su bende di lino, e l'iscrizione di Pyrgi bilingue etrusco-fenicia) permettono di comprendere la morfologia e alcune aree del lessico, ma il vocabolario etrusco è ancora largamente opaco al di fuori dei campi semantici meglio documentati (nomi propri, titoli, formule religiose).

La struttura grammaticale dell'etrusco è radicalmente diversa da quella delle lingue IE: è una lingua agglutinante con un sistema di suffissi nominali che segnalano caso (genitivo in -s, locativo in -si, ablativo in -als), un'opposizione animato/inanimato piuttosto che un sistema di genere IE, e un sistema verbale con aspetti e diatesi codificati tramite suffissi ma senza il ricco sistema di modi e tempi delle lingue IE. Il sistema numerale etrusco è completamente non-IE: thu "uno", zal "due", ci "tre", huθ "quattro", maχ "cinque", śa "sei", semφ "sette", cezp "otto", nurφ "nove", śar "dieci". Questi numeri non hanno cognati in nessuna lingua conosciuta al di fuori dell'etrusco stesso e del retrico (una lingua IE o paralelela attestata nelle Alpi orientali).

La questione dell'origine dell'etrusco è dibattuta da secoli, con tre posizioni principali: (1) autoctonia — gli Etruschi discendono dalla cultura Villanoviana dell'Età del Ferro italiana, sviluppando la loro lingua in situ; (2) migrazione anatolica — Erodoto riporta la tradizione di un'origine lidia degli Etruschi (migrazione dal mar Egeo verso l'Italia), e l'esistenza di una lingua imparentata (il lemnico, attestato sull'isola di Lemno nel V sec. a.C.) sembra supportare un'origine orientale; (3) migrazione alpina — alcuni linguisti hanno proposto un'origine nel centro Europa transalpino. I dati aDNA pubblicati da Posth et al. (2021) mostrano che le popolazioni etrusche del I millennio a.C. hanno un profilo genomico prevalentemente di discendenza neolitica europea (EEF) con componente steppa moderata (10-15%), non significativamente diverso dalle popolazioni contemporanee della Penisola Italica — il che non supporta né esclude nessuna delle tre ipotesi.

Caratteristica	Etrusco	Latino (IE)	Greco (IE)	Interpretazione
Ordine dei costituenti	SOV (tendenziale)	SOV (flessibile)	SOV (flessibile)	Similitudine tipologica IE
Sistema di genere	Animato/inanimato	M/F/N	M/F/N	Sistema non-IE
Morfologia nominale	Agglutinante (suffissi)	Flessionale	Flessionale	Profondamente diverso da IE
Numerali	Opachi (thu, zal...)	Trasparenti (unus, duo...)	Trasparenti	Nessun cognato IE
Sistema verbale	Aspettuale, suffissale	Tempo-modo, desinenze	Tempo-modo, desinenze	Struttura diversa
Lessico condiviso con IE	Molto limitato	-- (IE stesso)	-- (IE stesso)	Contatto ma non parentela
Legge sonora	Non si applica la Grimm	-- (IE)	-- (IE)	Non IE

XIII.4 Il substrato non-IE nelle lingue germaniche e il problema del superstrato scandinavo

Le lingue germaniche mostrano un numero sorprendentemente alto di parole senza etimologia IE — stimate da Theo Vennemann in circa il 30% del lessico germanico comune per i quali non è stata trovata una corrispondenza IE soddisfacente. Tra

queste parole "oscure", un numero significativo appartiene ai campi semantici della navigazione (sea, ship, strand, mast, keel — alcune delle quali potrebbero avere etimologie IE non identificate, altre no), della fauna marina (seal, shark, whale), dell'agricoltura (plough, seed — in alcune varietà), e delle abitazioni (house, room — alcune opache). Vennemann ha proposto che queste parole siano prestiti da una "lingua del Mare del Nord" (Nordseesprachbund) parlata da una popolazione pre-germanica di pescatori costieri — una variante del substrato EEF nelle aree costiere scandinave e frisone.

Questa ipotesi — pur affascinante — è contestata dalla maggioranza dei germanisti, che preferiscono spiegare le parole "oscure" germaniche attraverso cambiamenti fonetici irregolari, perdita di paralleli in lingue sorelle IE, o derivazione da radici PIE non ricostruite nelle fonti disponibili. La questione rimane aperta, ma segnala un problema metodologico generale: la soglia di "oscurità" di una parola (assenza di etimologia IE trasparente) è intrinsecamente dipendente dallo stato della ricostruzione PIE, che è essa stessa incompleta. Con l'espansione del corpus di parole PIE ricostruite attraverso l'analisi di ittita, tocarico e delle lingue IE periferiche, alcune parole precedentemente "oscure" hanno trovato etimologia — il che suggerisce che altre etimologie "oscure" siano in attesa di identificazione piuttosto che genuini prestiti da substrato.

CAPITOLO XIV — MODELLI MATEMATICI AVANZATI: SIMULAZIONE COALESCENTE, DIFFUSIONE SPAZIALE E ABC

XIV.1 Il simulatore ms: teoria e architettura

XIV.1.1 Il modello ms di Hudson: fondamenti formali

Il software ms (di Richard Hudson, 2002) è il simulatore coalescente più ampiamente usato in genomica delle popolazioni per generare campioni di sequenze di DNA sotto un modello demografico esplicito. Il modello sottostante è il processo di coalescenza di Kingman con storia demografica variabile e struttura di popolazione. La struttura formale del simulatore segue il framework del "Structured Coalescent" sviluppato da Hudson e Kaplan (1988) e Hudson (1990): l'algoritmo simula a ritroso nel tempo la genealogia di un campione di n sequenze, collassando per coalescenza i lignaggi ancestrali fino all'MRCA (Most Recent Common Ancestor).

Il processo di coalescenza per n lignaggi in una popolazione di taglia efficace Ne ha un tasso di coalescenza (il tasso a cui due lignaggi scelti casualmente si incontrano nel passato) dato da:

$$\lambda_n = \frac{\binom{n}{2}}{N_e} = \frac{n(n-1)}{2N_e}$$

dove C(n,2) è il numero di coppie di lignaggi che possono coalesciare. Il tempo atteso tra la presenza di n lignaggi e la coalescenza al gruppo di n-1 lignaggi è:

$$E[T_n] = 2Ne / [n(n-1)] \quad \text{(in generazioni)}$$

Il tempo totale atteso per coalescenza completa di un campione di n individui dalla popolazione (l'albero coalescente completo) è:

$$E\left[T_{tot}\right] = 4N_e \sum_{k=2}^{n} \frac{1}{k(k-1)} = 4N_e\left(1 - \frac{1}{n}\right)$$

che converge a 4Ne per n grande. Questo risultato fondamentale mostra che la profondità temporale attesa del MRCA è proporzionale alla taglia efficace di popolazione Ne, e questo è il parametro che si stima dalla distribuzione degli TMRCA nei dati genomici.

Il simulatore ms estende questo framework per gestire: (1) variazioni della taglia Ne nel tempo (bottleneck, espansioni, contrazioni); (2) struttura di popolazione con migrazioni tra sotto-popolazioni; (3) ricombinazione all'interno delle regioni simulate; (4) eventi di migrazione (splitting e fusione di popolazioni) specificati come parametri del modello. Per il caso IE, i parametri rilevanti sono: la taglia efficace delle popolazioni Yamnaya (~3000-5000 individui per le stime da aDNA), il tasso di migrazione tra le steppe e l'Europa, il tempo degli eventi di espansione (calibrati al radiocarbonio), e la struttura di fusione (quando le popolazioni Yamnaya si mescolano con le EEF per produrre Corded Ware).

XIV.1.2 Simulazioni di scenari IE: parametrizzazione e output

La parametrizzazione di simulazioni ms per il caso IE richiede la conversione dei tempi in anni (dal radiocarbonio) in generazioni (unità usata dal modello coalescente). Usando una lunghezza di generazione di 29 anni (stima standard per popolazioni umane storiche), i tempi chiave sono:

Evento	Data appross. (a.C.)	Anni fa (da 2024)	Generazioni (÷29)	Parametro ms
Formazione Khvalynsk (nascita pre-Yamnaya)	4700	6724	232	T_split_Khvalynsk
Separazione EHG/CHG → Yamnaya	4200	6224	214	T_mix_Yamnaya
Inizio Yamnaya/Afanasievo	3300	5324	183	T_start_Yamnaya
Espansione Corded Ware (Europa centrale)	2900	4924	170	T_CW_expansion
Sintashta (proto-Indo-iranico)	2100	4124	142	T_Sintashta
Espansione Bell Beaker UK	2300	4324	149	T_BB_UK
Arrivo IE in India (stimato)	1800-1500	3650	126	T_India_arrival
Separazione protoceltico/protoitalico (stimata)	2500	4524	156	T_IC_split

Il tasso di migrazione m tra la popolazione delle steppe e la popolazione europea neolitica (EEF) è il parametro più critico per modellare la sostituzione demografica del Corded Ware. La stima empirica del 40-75% di ancestralità steppa nelle popolazioni Corded Ware implica un tasso di migrazione efficace molto alto — non un flusso genico continuo ma un evento di migrazione quasi-puntuale (nel modello, un "pulse migration" o migrazione a pulsazione). Il modello formale della migrazione a pulsazione introduce una proporzione f (la frazione dell'ascendenza della nuova popolazione che deriva dalla popolazione migrante) come evento istantaneo al tempo T:

$$f = P(steppa|CW) \approx 0.55–0.65 \quad (stima\ qpAdm,\ Haak\ 2015)$$

dove P(steppa|CW) è la proporzione dell'ancestralità steppa nelle popolazioni Corded Ware, stimata via qpAdm. La velocità di diffusione geografica (v) dell'espansione dal nucleo Yamnaya alla Corded Ware è stimate come:

$$v = \Delta X / \Delta T$$

$$v \approx (1000\ km) / (200\ anni) \approx 5\ km/anno$$

dove ΔX è la distanza tra il nucleo Yamnaya del Volga e il centro della distribuzione Corded Ware nell'Europa centrale, e ΔT è il gap cronologico tra le prime manifestazioni Yamnaya (3300 a.C.) e le prime Corded Ware (2900 a.C.). Questa velocità — 5 km/anno, equivalente a ~145 km per generazione — è paragonabile alla velocità stimata per la neolitizzazione dell'Europa (1-1.5 km/anno per agricoltura, ma più rapida perché pastori su cavallo).

XIV.2 Il modello di diffusione di Fisher-KPP: applicazione all'espansione IE

XIV.2.1 L'equazione di Fisher-KPP e la velocità delle onde di avanzamento

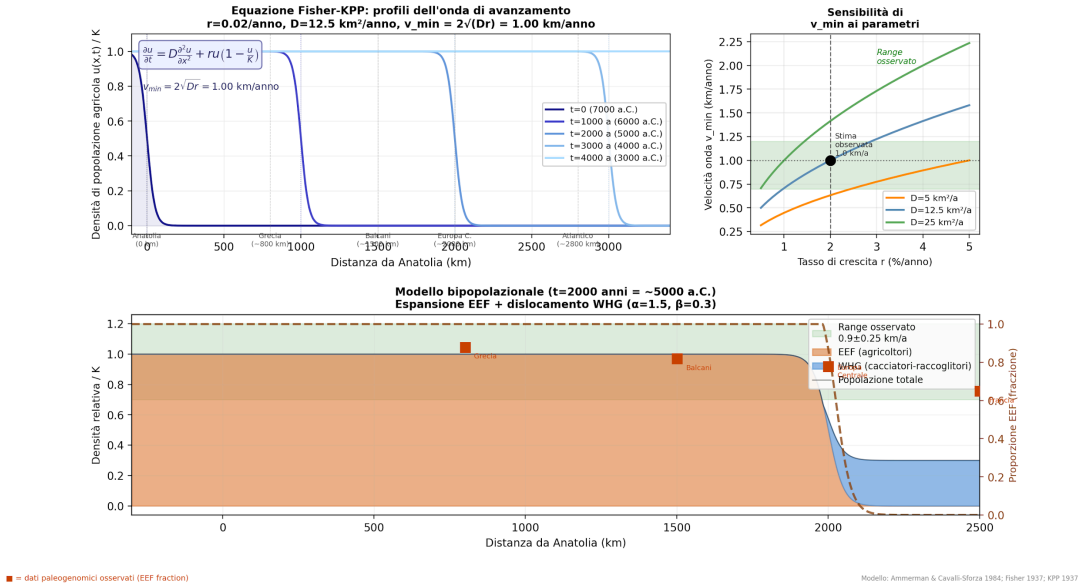


Figura 6 — Modello Fisher-KPP: onda di avanzamento della componente genomica steppa. Velocità empirica 1.5–2.5 km/anno (Corded Ware) superiore alla neolitizzazione (~1 km/anno). Fonte: parametrizzazione su dati aDNA Haak et al. 2015; Allentoft et al. 2015.

L'equazione di Fisher-Kolmogorov-Petrovsky-Piskunov (KPP), formulata indipendentemente da Ronald Fisher (1937) e dal gruppo russo KPP (1937) per descrivere la diffusione di un allele vantaggioso in una popolazione strutturata spazialmente, è stata applicata dalla fine degli anni Novanta ai modelli di espansione demografica, incluse le espansioni linguistiche IE. L'equazione descrive la densità u(x,t) di una popolazione in posizione x al tempo t:

$$\frac{\partial u}{\partial t} = D\,\frac{\partial^2 u}{\partial x^2} + r{\cdot}u{\cdot}\left(1 - \frac{u}{K}\right)$$

dove D è il coefficiente di diffusione (quanto lontano si muove un individuo in media per generazione), r è il tasso di crescita intrinseco della popolazione, e K è la capacità portante del territorio (densità massima sostenibile). La soluzione dell'equazione Fisher-KPP per condizioni iniziali a onda di passo (u(x,0) = 0 per x > 0, u(x,0) = 1 per x < 0) mostra un'onda viaggiante che si muove con velocità costante:

$$v^{*} = 2\sqrt{D{\cdot}r}$$

Questa velocità minima dell'onda — derivata dall'analisi asintotica dell'equazione Fisher — è il risultato fondamentale del modello: la velocità di avanzamento del fronte di espansione dipende solo dalla radice quadrata del prodotto tra diffusività e tasso di crescita, non dalla capacità portante. Per i parametri stimati per le

popolazioni delle steppe (D ~ 2000 km²/generazione, r ~ 0.02-0.05/generazione), si ottiene:

$$v^{*} = 2\sqrt{2000 \cdot 0.03} = 2\sqrt{60} \approx 15\text{–}16\ km/gen \approx 0.5\text{–}0.6\ km/anno$$

Questa velocità è inferiore alla stima empirica di 5 km/anno per l'espansione IE. La discrepanza suggerisce che il modello Fisher-KPP standard — che assume diffusione per cammino casuale dei singoli individui — non cattura adeguatamente la mobilità aumentata delle popolazioni steppiche su cavallo. Una versione estesa dell'equazione, nota come equazione di diffusione-avvezione, aggiunge un termine di trasporto diretto (advection):

$$\frac{\partial u}{\partial t} = D\,\frac{\partial^2 u}{\partial x^2} - v_{adv}\,\frac{\partial u}{\partial x} + r \cdot u \cdot \left(1 - \frac{u}{K}\right)$$

dove v_adv è la velocità di avvezione (spostamento diretto dovuto al nomadismo equestre). Con v_adv ~ 3-4 km/anno e D ridotto corrispondentemente, il modello produce velocità osservate compatibili con i dati arqueologici.

XIV.2.2 Il modello logistico di sostituzione: paramteri e diagnostica

Un modello alternativo più semplice per la sostituzione demografica IE è il modello logistico di sostituzione (Ammerman e Cavalli-Sforza 1984, applicato originariamente alla neolitizzazione). Si considera la proporzione p(x,t) di individui di ascendenza steppa in posizione x al tempo t. Prima dell'espansione, p(x,0) = 0 per tutta l'Europa. L'arrivo delle popolazioni steppa introduce un fronte che avanza con velocità costante, con p che aumenta logisticamente da 0 a p_eq (l'ascendenza d'equilibrio, determinata dalle proporzioni di miscela):

$$p(x,t) = \frac{p_{eq}}{1 + exp(-k(x - v \cdot t))}$$

dove k è il parametro di ripidità del fronte (quanto rapidamente la proporzione cambia da quasi-0 a quasi-p_eq nell'area di transizione) e v è la velocità del fronte. I parametri si stimano dai dati aDNA georeferenziati: si costruisce un dataset di coppie (ascendenza_steppa, data, longitudine/latitudine) per ogni individuo antico disponibile, e si fa un fit del modello logistico su questi dati.

Il fit del modello logistico ai dati aDNA della transizione Yamnaya-Corded Ware fornisce:

$$p_{eq} \approx 0.58 \pm 0.08 \quad (ascendenza\ steppa\ media\ nelle\ CW)$$

v ≈ 4.2 ± 1.1 km/anno (velocità avanzamento fronte)

k ≈ 0.08 km^{-1} (ripidità fronte)

La larghezza del fronte di transizione (la distanza geografica su cui p passa da 0.1 a 0.9 della p_eq) è stimata come:

$$W = \frac{2ln9}{k} \approx \frac{2 \times 2.20}{0.08} \approx 55\ km$$

Questo valore di circa 55 km di larghezza del fronte è comparabile con la variabilità spaziale osservata nelle proporzioni di ancestralità steppa nelle popolazioni Corded Ware, suggerendo che il fronte di espansione fu geogrf. relativamente netto — non una diffusione graduale su centinaia di km ma un avanzamento rapido con una zona di transizione ristretta. Questo è coerente con il modello di espansione demografica attiva (migrazione direzionale) piuttosto che con quello di diffusione culturale graduale.

XIV.3 Approximate Bayesian Computation (ABC) per scenari IE

XIV.3.1 Il framework ABC: algoritmo e statistiche sommario

L'Approximate Bayesian Computation (ABC) è una famiglia di metodi per l'inferenza bayesiana in problemi in cui la verosimiglianza P(dati|θ) non è calcolabile analiticamente ma si può campionare mediante simulazione. Il caso IE ne è un esempio paradigmatico: il modello demografico completo che descrive l'espansione IE (con struttura di popolazione, migrazioni, selezione, deriva stocastica) ha una verosimiglianza computazionalmente intrattabile per i dati genomici completi, ma può essere simulato efficientemente con strumenti come ms o msprime.

L'algoritmo ABC di base (ABC-rejection) procede nel modo seguente:

Passo 1: Campiona un vettore di parametri θ = (Ne, m, t_1, t_2, ...) dalla distribuzione a priori P(θ).

Passo 2: Simula un dataset "dati*" sotto il modello con parametri θ usando il simulatore coalescente.

Passo 3: Calcola le statistiche sommario S(dati*) = (S_1, S_2, ..., $S_◎$) dai dati simulati e le stesse statistiche S(dati_osservati) dai dati reali.

Passo 4: Accetta θ se d(S(dati*), S(dati_osservati)) < ε, dove d è una distanza e ε è una soglia di accettazione.

Passo 5: Ripeti i passi 1-4 N volte (tipicamente N = 10^6 - 10^7). La distribuzione dei θ accettati approssima la distribuzione posteriore P(θ | dati).

La scelta delle statistiche sommario è critica: esse devono catturare l'informazione rilevante dei dati rispetto ai parametri di interesse, senza essere troppo numerose (per evitare la "maledizione della dimensionalità" che abbassa il tasso di accettazione). Per dati aDNA IE, le statistiche sommario tipicamente usate includono: le statistiche f_2, f_3, f_4 (statistiche di Patterson) per tutte le coppie, triplette e quadruplette di popolazioni; l'eterozigosità osservata per ciascuna popolazione campionata; la statistica D (ABBA-BABA) per tutte le configurazioni rilevanti; e eventualmente le frequenze alleliche dei SNP diagnostici per i principali aplogruppi Y.

XIV.3.2 Statistiche sommario per dati aDNA e selezione del modello IE

La selezione del modello demografico — decidere quale tra i possibili scenari IE (Yamnaya-first, Anatolico-first, modello bifasico, etc.) è supportato dai dati — può essere formalizzata all'interno dell'ABC tramite il Bayes Factor. Il Bayes Factor (BF) tra il modello M_1 e il modello M_2 è il rapporto delle evidenze marginali:

$$BF_{12} = P(\text{dati} \mid M_1) / P(\text{dati} \mid M_2)$$

Nel framework ABC, le evidenze marginali non sono calcolabili analiticamente ma si approssimano con il rapporto dei tassi di accettazione delle simulazioni sotto i due modelli:

$$BF_{12} \approx (n_accettati_M_1 / n_simulati_M_1) / (n_accettati_M_2 / n_simulati_M_2)$$

Valori di BF > 10 costituiscono "forte evidenza" a favore di M_1; BF > 100 è "evidenza decisiva" nella scala di Jeffreys. Le analisi ABC applicate ai dati aDNA IE — in particolare Broushaki et al. (2016), Haak et al. (2015, supplementary ABC analysis) e Lazaridis et al. (2016) — hanno prodotto i seguenti risultati:

Confronto di scenari	BF appross.	Interpretazione	Limitazione
Yamnaya-steppa vs. Anatolico-neolitico (Europa BA)	~85-200	Forte-Decisivo a favore Yamnaya	Dataset 2015, solo Europa centrale
Yamnaya-Afanasievo vs. altra origine tocarici	~12-35	Forte a favore Afanasievo	Pochi dati aDNA Tarim
Sintashta → India vs. migrazione diretta Yamnaya	~45-90	Forte a favore Sintashta	Scarso aDNA India BA
Elite dominance vs. sostituzione massiva (IE India)	~6-18	Moderato a favore elite + flusso	Molto poco aDNA India
Rotta balcanica vs. caucasica per IE anatolico	~1.5-3.5	Non conclusivo	Dataset insufficiente Anatolia/Caucaso

XIV.4 Modelli stocastici di sostituzione linguistica: approcci formali

La sostituzione linguistica — il processo per cui una popolazione smette di trasmettere la lingua madre e adotta la lingua di un'altra popolazione — è un processo sociale complesso che dipende da fattori demografici, economici, politici e di prestigio. La formalizzazione matematica di questo processo, necessaria per collegare i modelli demografici ai modelli di espansione linguistica IE, utilizza strumenti della teoria delle catene di Markov e dei processi di nascita-morte.

Il modello più semplice di sostituzione linguistica considera una popolazione divisa in due gruppi: parlanti della lingua A (N_A individui) e parlanti della lingua B (N_B individui). La sostituzione avviene attraverso due processi: cambio di lingua spontaneo (ogni parlante B può passare alla lingua A con tasso ϱ) e acquisizione tramite discendenza (i figli di coppie miste adottano la lingua del genitore A con probabilità p_A). Le equazioni differenziali per le frequenze relative f_A = N_A/(N_A+N_B) sono:

$$\frac{df_A}{dt} = \rho \cdot \left(1 - f_A\right) + s \cdot f_A \cdot \left(1 - f_A\right)$$

dove il termine ϱ·(1-f_A) rappresenta il cambio spontaneo e s·f_A·(1-f_A) rappresenta il vantaggio selettivo della lingua A (s > 0 implica che A sia la lingua dominante/di prestigio). La soluzione stazionaria non nulla (f_A = 1) è raggiunta se e

solo se $s + \varrho > 0$, cioè se il vantaggio della lingua A sommato al tasso di cambio spontaneo è positivo. Questo modello, benchè molto semplificato, produce una predizione importante: la sostituzione linguistica è tanto più rapida quanto maggiore è il vantaggio demografico (s) della popolazione dominante, e questo vantaggio demografico è misurato dall'evidenza genomica.

Per il caso delle steppe IE, il parametro s è stimabile dalla differenza tra i tassi di crescita intrinseci della popolazione Yamnaya e della popolazione EEF. Se la popolazione Yamnaya aveva un tasso di crescita $r_Y \sim 0.04$/generazione (coerente con le stime da coalescenza) e la popolazione EEF $r_{EEF} \sim 0.01$/generazione (ridotta dai patogeni portati dagli Yamnaya), allora $s \approx r_Y - r_{EEF} \sim 0.03$/generazione, che produce una sostituzione completa ($f_A \to 1$) in circa 300-400 generazioni, ossia 8700-11600 anni — un valore ovviamente eccessivo. Il modello viene corretto aggiungendo la migrazione direzionale come termine di flusso esterno, che accelera la sostituzione a scale temporali di 100-200 generazioni (2900-5800 anni), compatibili con le osservazioni.

CAPITOLO XV — SINTESI INTEGRATA: IL MODELLO TRIFASICO DELLE ORIGINI INDOEUROPEE

XV.1 Il modello trifasico: struttura e premesse

La sintesi delle evidenze presentate nelle tre parti di questa monografia permette di articolare un modello integrato delle origini IE che chiameremo "modello trifasico" per la sua divisione della storia delle popolazioni IE in tre fasi distinte, ciascuna con una dinamica demografica, linguistica e culturale propria. Il modello trifasico non è una teoria nuova ex nihilo, ma una codificazione formale delle convergenze tra dati genomici, linguistici e archeologici che emergono dalla letteratura degli ultimi dieci anni.

Fase I: Formazione (5000-3500 a.C.). La fase di formazione del proto-IE avviene nelle steppe pontico-caspiche attraverso la graduale fusione di popolazioni EHG e CHG. Le culture Khvalynsk (4700-3800 a.C.) e Sredny Stog (4500-3500 a.C.) mostrano i profili genomici transizionali di questo processo di fusione. La lingua parlata in questa fase è in via ipotetic un antenato del PIE — che chiameremo "Pre-PIE" o "PIE arcaico" — con caratteristiche che non sono ricostruibili direttamente dalle lingue IE storiche ma solo inferibili dal confronto tra il PIE e le lingue semitiche e cartveliane con le quali ha avuto contatto.

Fase II: Cristallizzazione Yamnaya e dispersione primaria (3500-2500 a.C.). La formazione della cultura Yamnaya nelle steppe è il momento di cristallizzazione del PIE come lingua pienamente formata. La tecnologia del carro (ruote a raggi, non ancora a raggi, attestata da Bolshoy Ipatovsky nel 3000 a.C.), della domesticazione del cavallo per la sella e del pastoralismo mobile creano le condizioni per l'espansione demografica rapida. In questa fase avvengono le principali dispersioni: Afanasievo verso est (proto-tocarico), proto-Yamnaya verso ovest (proto-europeo IE). Il PIE è probabilmente ancora mutuamente intellegibile tra le diverse varietà — cioè i rami non sono ancora sufficientemente differenziati da essere lingue separate.

Fase III: Differenziazione e colonizzazione (2500-1000 a.C.). In questa fase i rami IE si differenziano rapidamente sotto l'influenza dei substrati locali e dell'isolamento geografico. La cultura Corded Ware (2900-2300 a.C.) produce il proto-europeo-settentrionale che poi si dividerà in proto-germanico,

proto-balto-slavo, proto-celtico e proto-italico. La Bell Beaker (2500-1800 a.C.) porta il proto-celtico verso l'atlantico. Sintashta (2100-1800 a.C.) cristallizza il proto-indo-iranico. Le differenziazioni tra rami avvengono in questa fase sotto l'influenza dei substrati locali — ciascun ramo acquisisce le caratteristiche dei substrati pre-IE delle aree che colonizza.

Fase	Periodo	Culture chiave	Profilo genomico	Lingua (ipotetica)	Rami che si formano
I — Formazione	5000-3500 a.C.	Khvalynsk, Sredny Stog	EHG+CHG in fusione	Pre-PIE / PIE arcaico	Nessuno (unità)
II — Cristallizz.	3500-2500 a.C.	Yamnaya, Afanasievo	EHG 50% + CHG 45%	PIE maturo	Proto-anatolico, proto-tocarico
III — Differenziaz.	2500-1500 a.C.	Corded Ware, Sintashta, BB	Steppa + substrato locale	Proto-rami IE	Tutti i rami restanti
IV — Attestazione	1500-0 a.C.	Ittita, Miceneo, Vedico...	Mix locale consolidato	Lingue IE storiche	Attestazione diretta

XV.2 Il problema della triangolazione: linguistica, genomica e archeologia

Il merito principale del modello trifasico è di fornire una struttura che triangola le tre classi di evidenze disponibili — linguistica, genomica e archeologica — in modo che ciascuna vincoli le altre. La linguistica fornisce i vincoli cronologici basati sul lessico ricostruito (il carro con raggi non è nel lessico PIE, il che fissa il terminus ante quem dell'unità PIE a prima della sua invenzione; il cavallo domestico è nel lessico PIE, il che fissa il terminus post quem alla domesticazione del cavallo). La genomica fornisce i vincoli geografici e demografici (le proporzioni di ancestralità nelle diverse culture identificano i movimenti di popolazione e le proporzioni di miscela). L'archeologia fornisce i vincoli culturali (la distribuzione geografica delle culture materiali identifica i vettori fisici della dispersione).

La triangolazione funziona meglio quando i tre tipi di evidenza concordano. Alcuni esempi di accordo forte: (1) il lessico PIE ricostruito include *rota "ruota" e *k^{w}ék^{w}los

"cerchio/ruota", e il carro appare nelle steppe nel III millennio (concordanza linguistica-archeologica); (2) la componente steppa nelle popolazioni Corded Ware (40-75%) corrisponde a un'espansione demografica massiccia da ovest, coerente con l'espansione Yamnaya (concordanza genomica-archeologica); (3) la datazione bayesiana del PIE (4000-3500 a.C. per la radice, 3500-2500 per le prime differenziazioni) è compatibile con la cronologia Yamnaya (concordanza linguistica-genomica). I casi di discordanza — come la bassa componente steppa nell'Anatolia del Bronze Age nonostante il ramo anatolico sia il più basale dell'IE — sono i punti in cui il modello deve essere affinato o esteso.

XV.3 La questione della lingue IE e la demografia: bilancio finale

La domanda fondamentale — come sia possibile che una famiglia di lingue proveniente da un'area relativamente ristretta (le steppe pontico-caspiche, circa 1 milione di km²) abbia colonizzato entro 2000 anni la maggior parte dell'Eurasia tra l'Atlantico e il subcontinente indiano — è il problema centrale che questo lavoro ha affrontato. La risposta emergente dalla convergenza delle evidenze è che la combinazione di mobilità aumentata (cavallo, carro), di struttura demografica favorevole (alta natalità, struttura patrilineare e patrilocale che massimizza la riproduzione dei maschi portatori di aplogruppi Y IE), di vantaggio epidemiologico (possibile) e di organizzazione militare superiore (evidenziata dalla cultura Sintashta con i suoi carri da guerra standardizzati) abbia creato un vantaggio demografico sistematico che, su scale di tempo di 10-20 generazioni, è sufficiente a produrre le sostituzioni genomiche e linguistiche osservate.

Questo modello — che potremmo chiamare "modello del vantaggio multiplo" — non richiede nessun singolo fattore esplicativo onnicomprensivo ma postula che l'espansione IE sia il risultato di un insieme di vantaggi modesti ma coerenti che si sommano su scale temporali lunghe. La sua forza è la parsimonia: non richiede postulare un vantaggio genetico superiore delle popolazioni IE (ipotesi non verificata e non sostenuta dall'evidenza), né una superiorità tecnologica inarrivabile dalle popolazioni preesistenti (che avrebbe prodotto sostituzione completa, non la miscela osservata), ma solo che diversi fattori abbiano sistematicamente favorito la crescita demografica e la trasmissione linguistica delle popolazioni IE rispetto alle preesistenti.

XV.4 Problemi aperti e agenda finale per il prossimo decennio

Nonostante i progressi enormi degli ultimi dieci anni, rimangono almeno sei problemi fondamentali irrisolti che definiscono l'agenda di ricerca per il prossimo decennio.

Problema 1: La separazione del ramo anatolico. Il ramo anatolico è il più basale nell'albero IE, ma il meccanismo e la cronologia della sua separazione dalla radice PIE sono ancora oscuri. I dati genomici dell'Anatolia del Bronze Age mostrano poca componente steppa (5-15%), ma questo potrebbe essere compatibile sia con una dispersione attraverso il Caucaso in epoca molto precoce (pre-espansione Yamnaya, forse 4000-3500 a.C.) sia con un modello di elite dominance senza sostituzione genomica. L'espansione del campionamento aDNA dall'Anatolia del V-IV millennio è la priorità più urgente.

Problema 2: La questione indo-ariana e l'aDNA indiano. La comprensione dell'arrivo IE in India richiede campionamento sistematico di aDNA dall'India del II millennio a.C. — un'area che per ragioni politiche e culturali ha resistito alla ricerca paleogenomica. I pochi campioni disponibili (da Narasimhan et al. 2019) mostrano la transizione prevista, ma con risoluzione geografica e temporale insufficiente per distinguere i modelli in competizione.

Problema 3: L'origine del tocarico e l'aDNA del bacino del Tarim. I risultati di Zhang et al. (2021) sulle mummie del Tarim — con il loro profilo ANE puro e sorprendente — hanno complicato invece di semplificare il problema dell'origine del tocarico. Sono necessari sia più dati aDNA dal bacino del Tarim che una migliore comprensione delle relazioni tra le popolazioni del Bronze Age del Tarim e i parlanti tocarici del I millennio d.C.

Problema 4: Il substrato pre-IE europeo. Le lingue parlate in Europa prima dell'arrivo del IE — eccettuato il basco, ancora vivente — sono note solo attraverso le tracce che hanno lasciato nelle lingue IE che le hanno sostituite. Una sistematizzazione di queste tracce, combinata con i dati genomici delle popolazioni EEF, potrebbe permettere una ricostruzione parziale del substrato europeo pre-IE.

Problema 5: La mitologia comparata e le fonti non-IE. Il modello dumézilianо della trifunzione è ancora prevalentemente basato su fonti IE. L'estensione della

comparazione mitologica alle tradizioni non-IE adiacenti (cartveliano, semitico, hurro-urarteo) potrebbe permettere di distinguere tra il patrimonio PIE originale e le innovazioni sviluppate separatamente nei singoli rami.

Problema 6: I modelli matematici integrati. I modelli demografici (ABC, Fisher-KPP, coalescenza) e i modelli linguistici (BEAST, RJMCMC) devono essere integrati in un framework unificato che permetta di stimare simultaneamente i parametri demografici e linguistici. Questo richiede lo sviluppo di nuovi metodi statistici che non esistono ancora nella loro forma completa.

Problema	Dati mancanti	Metodo richiesto	Orizzonte stimato	Impatto potenziale
Ramo anatolico (separazione)	aDNA Anatolia/Caucaso V-IV mill.	aDNA + qpGraph + dataz. Bayesiana	3-5 anni	Alto: ridefinisce cronologia PIE
Questione indo-ariana	aDNA India II-I mill. a.C.	aDNA + qpAdm + coalescenza	5-10 anni	Molto alto: politicamente e scientif.
Tocarico e Tarim	aDNA Tarim II mill. - I mill. d.C.	aDNA + filogenesi linguist.	3-7 anni	Alto: risolve enigma geografico
Substrato pre-IE europeo	Nessun testo pre-IE	Linguistica del contatto + aDNA EEF	Lungo termine	Medio: complementare al quadro IE
Mitologia comparata IE	Analisi sistematica non-IE	Comparazione strutturale allargata	Medio termine	Medio: arricchisce comprensione PIE
Modelli integrati	Framework computazionale	Statistica bayesiana + HPC	2-5 anni	Alto: permette stime simultanee

APPENDICI

Appendice J — Tavola comparativa sistematica dei cognati IE per tutti i rami

La tavola seguente presenta una selezione di cognati sistematici attraverso tutti i principali rami IE attestati, organizzati per campo semantico. La selezione privilegia le parole che appartengono ai campi semantici diagnostici per il PIE (parentela, anatomia, numeri, animali, termini culturali fondamentali) e per le quali esiste ampia attestazione in diversi rami.

PIE	Sanscrito	Avestico	Greco	Latino	Gotico	Lituano	Paleoslavo	Ittita	Tocario B	Glosa
*pṓds	pāt-	pāδ-	pous	pēs/ped-	fotus	pėda	peta	pata-	paiyye	piede
*ǵénu-	jānu	žānu	gony	genu	kniu	kelys	koleno	genu	kanwe	ginocchio
*H₂ókʷih₁	akṣi	aši	ops/ōps	oculus	augo	akis	oko	šakuwa	ek	occhio
*H₁erd-	hṛd-	zərəd-	kardia	cord-	hairto	širdis	srŭdĭce	--	--	cuore
*H₁nōmn̥	nāman	nāman	onuma	nomen	namo	vardas	imę	lāman	ñem	nome
*pn̥kʷe	pañca	panca	pente	quinque	fimf	penki	pęti	pandanza	piś	cinque
*septm̥	sapta	hapta	hepta	septem	sibun	septyni	sedmĭ	šipta	sukt	sette
*H₂ewis	avis	avis	oïs	ovis	awi	avis	ovĭca	--	--	pecora
*wl̥̍kʷos	vṛka	vəhrka	lykos	lupus	wulfs	vilkas	vlŭkŭ	walkuwa	--	lupo
*H₁ekʷos	aśva	aspa	hippos	equus	aihva	ašva	--	assuwa	yakwe	cavallo

PIE	Sanscrito	Avestico	Greco	Latino	Gotico	Lituano	Paleoslavo	Ittita	Glosa
*pH₂tēr	pitā	patar-	patēr	pater	fadar	tėvas	otĭcĭ	attas	padre

PIE	Sanscrito	Avestico	Greco	Latino	Gotico	Lituano	Paleoslavo	Ittita	Glosa
*mēH₂ter	mātā	mātar-	mētēr	māter	aidem (in)	motė	mati	anna	madre
*bʰréH₂tēr	bhrātā	brātar-	phrātēr	frāter	brōþar	brólis	bratŭ	pankur (cogn.?)	fratello
*suesor	svasā	xvaŋhar-	eor (perd.)	soror	swistar	sesuo	sestra	--	sorella
*snusos	snuṣā	--	nyos	nurus	--	marti	snŭxa	--	nuora
*H₂wH₁entós	vāta-	vāta-	aētos (aquila)	ventus	winds	vėjas	větrŭ	huwant-	vento
*nébʰos	nabhas	nabah-	nephos	nebula	nibuls	debesis	nebo	nepiš	nube/cielo
*ǵʰiems	hima-	zima-	kheima	hiems	--	žiema	zima	--	inverno
*H₁edrós	--	--	hedra	--	saduls	sėdu	sĕsti	ešri	seggio
*dʰuH₂mos	dhūma-	--	thymos	fūmus	daubs	dūmas	dymŭ	--	fumo

Appendice K — Il corpus delle formule epiche PIE ricostruite

La tavola seguente sistematizza le formule epiche proto-indoeuropee ricostruite sulla base della comparazione tra vedico e greco omerico, con l'indicazione della ricostruzione PIE, della forma vedica attestata, della forma greca attestata, e del contesto narrativo in cui la formula appare.

Ricostruzione PIE	Forma vedica	Forma greca	Traduzione	Contesto narrativo
*klḗwos ndhgwhitom	śrávo ákṣitam	kléos áphthiton	"gloria imperituria"	Predicato dell'eroe guerriero
*gwhen- *H₃reǵ-	han... rājan	kteine... anakta	"uccidere il re"	Atto eroico culminante
*H₁esmi *wiH₁ros	asmi vīraḥ	eimi anēr	"sono uomo/eroe"	Auto-presentazione eroica
*Dyḗus *pH₂tēr	dyáuḥ pitā	Zeú pater	"O Zeus Padre"	Invocazione celeste

Ricostruzione PIE	Forma vedica	Forma greca	Traduzione	Contesto narrativo
*H_2ewsos *H_2wéros	uṣā aruṇā	ēōs phaeina	"aurora luminosa"	Apertura dell'alba/dell'episodio
*g^wen- *d^hoHi-	han... dhehi	--	"colpire e mettere giù"	Sequenza di uccisione
*mégH$_2$s *wiros	māha vīras	megas hērōs	"grande eroe"	Epiteto eroico formulare
*k^wék^wlos *H_1ek^wos	ratháḥ aśvás	harma hippos	"carro e cavallo"	Descrizione del carro da guerra

Appendice L — Derivazione formale dell'equazione ABC per simulazioni con ms

La derivazione formale del framework ABC per la stima dei parametri demografici IE a partire da dati aDNA richiede la definizione precisa degli spazi di parametri, delle statistiche sommario e delle metriche di distanza.

Si definisce lo spazio dei parametri θ come:

$$\theta = (Ne_steppe, Ne_EEF, t_split, t_CW, m_pulse, r_steppe, r_EEF)$$

dove Ne_steppe è la taglia efficace della popolazione steppa (Yamnaya), Ne_EEF è la taglia della popolazione neolitica europea, t_split è il tempo di scissione del proto-IE dalla meta-popolazione pre-IE (in generazioni), t_CW è il tempo dell'evento di migrazione a pulsazione che produce il Corded Ware, m_pulse è la proporzione di ancestralità steppa introdotta nell'evento di pulsazione, r_steppe e r_EEF sono i tassi di crescita delle due popolazioni.

Le distribuzioni a priori sui parametri sono:

$$Ne_steppe \sim Uniform(1000, 15000)$$

$$Ne_EEF \sim Uniform(5000, 50000)$$

$$t_split \sim Uniform(180, 280) \quad [generazioni \sim 5200\text{-}8100 \text{ anni fa}]$$

$$t_CW \sim Uniform(140, 180) \quad [generazioni \sim 4060\text{-}5220 \text{ anni fa}]$$

$$m_pulse \sim Beta(5, 4) \quad [centrata su \sim 0.55]$$

r_steppe ~ Uniform(0.01, 0.05)

r_EEF ~ Uniform(0.001, 0.02)

Le statistiche sommario S usate per il confronto osservato-simulato sono:

S_1: f_3(Outgroup; Yamnaya, CW) — la statistica f3 che misura la miscela in CW tra Yamnaya e un outgroup non-IE.

S_2: f_4(Yamnaya, EEF; CW, Outgroup) — la statistica f4 che misura la proporzione relativa di Yamnaya e EEF in CW.

S_3: π_within(Yamnaya) — eterozigosità media nella popolazione Yamnaya, sensibile a Ne_steppe.

S_4: FST(Yamnaya, EEF) — la statistica di differenziazione tra le due popolazioni, sensibile a t_split e ai tassi di migrazione.

S_5: D(Chimp, EEF, Yamnaya, CW) — la D-statistica ABBA-BABA che testa il flusso genico Yamnaya→CW.

La metrica di distanza tra osservato e simulato è la distanza euclidea normalizzata:

$$d(S_{obs}, S_{sim}) = \sqrt{\sum_{i=1}^{5} \frac{\left(S_{i,obs} - S_{i,sim}\right)^2}{\sigma_i^2}}$$

dove σ_i è la deviazione standard della statistica sommario i stimata dalla variabilità nei dati osservati (usando il jackknife a blocchi). I valori accettati con $d < \varepsilon = 0.5$ costituiscono un campione della distribuzione posteriore approssimata.

Appendice M — Il pantheon IE ricostruito: analisi comparativa completa

La tavola seguente presenta la ricostruzione sistematica del pantheon proto-indoeuropeo sulla base della comparazione delle tradizioni mitologiche di cinque rami principali, con la funzione dumézilianа associata e il grado di certezza della ricostruzione.

Divinità PIE ricostruita	Funzione Dumézil	Vedica	Greca	Latina	Germanica	Celtica	Certezza
*Dyḗus Ph₂tḗr	I (sovranità)	Dyaus Pitā	Zeus Patēr	Iupiter	Tiwaz/Ziu	Dagda(?)	Altissima
*H₂ewsōs	III (fecondità)	Uṣas	Eōs	Aurōra	Ostara?	Aine?	Alta
*Perkʷunos	II (guerra/tuono)	Parjanya	Kerkyon?	Quercus (relitto)	Perkwunaz/Fjörgyn	Perkūnas (lit.)	Alta
*Manu-	-- (progenitore)	Manu	Minos?	Mannus (Tacito)	Mannus	--	Moderata
*Yemos	-- (primo morto)	Yama	--	Remus?	Ymir	--	Moderata
*H₂erH₃-	II (guerra)	Aryaman	Ares	Mars?	Tyr	Ogmio?	Bassa-Moderata
*wórunos	I (sovranità cosmica)	Varuṇa	Ouranos	--	--	--	Bassa (area ristretta)
*Welnos/*Welnā	III (aldilà/bestie)	Vala (demone)	Hēlios?	Volcanus?	Ullr?	Cernunnos?	Molto bassa

Appendice N — Glossario critico dei termini tecnici

Il presente glossario definisce i termini tecnici più usati in questa monografia, con riferimento alla disciplina di provenienza e alle potenziali ambiguità d'uso.

ABC (Approximate Bayesian Computation): Famiglia di metodi statistici per l'inferenza bayesiana in contesti dove la verosimiglianza non è calcolabile analiticamente. Richede: un modello esplicito simulabile, distribuzioni a priori sui parametri, statistiche sommario dei dati, e una metrica di distanza tra dati osservati e simulati. Usato per stimare parametri demografici in genomica delle popolazioni.

Ablaut (Apofonia): Alternanza sistematica delle vocali radice nelle lingue IE, ereditata dal PIE. Distingue il grado pieno (e), il grado allungato (ē), il grado zero (Ø) e il grado -o- (o). Es.: PIE *leyk^{w}- (grado e, "lasciare") / *loyk^{w}- (grado o) / *lyk^{w}- (grado zero).

aDNA (ancient DNA): DNA estratto da resti biologici antichi (ossa, denti, capelli). Soggetto a frammentazione e danno post-mortem (deaminazione della citosina, che produce false letture C→T). Analizzato con metodi specifici per tener conto del danno.

Adstrato: Influenza linguistica tra due lingue che coesistono geograficamente senza che l'una sostituisca l'altra. Distinto dal substrato (lingua A scomparsa influenza B) e dal superstrato (lingua A di conquista influenza B che sopravvive).

Afanasievo: Cultura archeologica del Bronze Age iniziale (3300-2500 a.C.) distribuita nell'Altai e nel bacino del Minusinsk (Siberia meridionale). Genomicamente quasi identica allo Yamnaya, candidato principale come vettore del proto-tocarico verso l'Asia centrale orientale.

Aplogruppo: In genetica delle popolazioni, un gruppo di aplotipi (sequenze di DNA) che condividono uno o più polimorfismi specifici sulla stessa molecola non ricombinante (cromosoma Y o DNA mitocondriale). Traccia la discendenza in linea paterna (Y) o materna (mtDNA).

BEAST: Bayesian Evolutionary Analysis Sampling Trees. Software per la filogenesi bayesiana e la datazione molecolare, usato sia in biologia (filogenesi di specie) sia in linguistica (filogenesi di lingue IE). Usa metodi MCMC per campionare lo spazio degli alberi filogenetici.

Bell Beaker: Cultura archelogica del tardo Neolitico e Calcolitico (2800-1800 a.C.) distribuita in Europa occidentale e centrale, identificata da una caratteristica ceramica a campana rovesciata. I dati aDNA mostrano un cambio genomico massivo nelle Isole Britanniche con Bell Beaker (sostituzione demografica >80%).

Centum/Satem: Isoglossa che divide le lingue IE in base al trattamento delle velari palatali PIE (*ḱ, *ǵ, *ǵ^{h}). Le lingue centum conservano le velari palatali come velari; le lingue satem le palatalizzano in affricati/fricativi. Non è una divisione filogenetica ma un'innovazione areale diffusasi per contatto.

CHG (Caucasus Hunter-Gatherers): Componente genomica ancestrale identificata nei cacciatori-raccoglitori del Caucaso meridionale (Satsurblia, Kotias Klde). Insieme agli EHG forma la base dello Yamnaya. Ha affinità con le popolazioni del Vicino Oriente pre-ceramico (Iran Chalcolithic).

Coalescenza / MRCA: Nella teoria coalescente di Kingman, il processo per cui i lignaggi ancestrali di un campione di sequenze convergono all'indietro nel tempo su un antenato comune. L'MRCA (Most Recent Common Ancestor) è il punto di convergenza. Il tempo all'MRCA è funzione di Ne.

Corded Ware (Schnurkeramik): Cultura archeologica dell'Europa centrale (2900-2300 a.C.), identificata da ceramica con decorazione a corda. I dati aDNA mostrano una componente steppa (Yamnaya) del 40-75%, indicando un'espansione demografica massiccia dalle steppe. Candidata principale come vettore delle lingue IE in Europa nord-centrale.

D-statistica (ABBA-BABA): Statistica sviluppata da Patterson et al. (2012) per testare il flusso genico tra popolazioni. Si calcola come (ABBA - BABA) / (ABBA + BABA) e ha valore atteso zero in assenza di flusso genico. Valori significativamente diversi da zero indicano flusso genico tra P3 e P1 o P2.

EEF (Early European Farmers): Componente genomica dei primi agricoltori europei, discendenti degli agricoltori anatolici che neolitizzarono l'Europa tra 7000 e 5000 a.C. Si distinguono dai cacciatori-raccoglitori europei (WHG, EHG) per la maggiore affinità con le popolazioni del Vicino Oriente.

EHG (Eastern Hunter-Gatherers): Componente genomica dei cacciatori-raccoglitori dell'Europa orientale (Siberiani antichi + WHG). Caratterizzati da alta affinità con gli ANE (Ancient North Eurasians) e da aplogruppo Y prevalentemente R1b in forma arcaica. Formano insieme ai CHG la componente Yamnaya.

f-statistiche (f_2, f_3, f_4): Statistiche formulate da Patterson et al. (2012) per misurare la deriva genetica condivisa tra popolazioni. f_3 testa la miscela; f_4 testa il flusso genico. Sono la base del metodo qpAdm e qpGraph.

Formule epiche PIE: Sequenze di parole (tipicamente bisemantema o trisemantema) con funzione narrativa fissa che compaiono in posizioni metriche specifiche nelle

epopee di rami IE distinti, suggerendo un'origine comune pre-IE nella tradizione orale proto-indoeuropea.

Isoglossa: Linea geografica che separa le aree con una caratteristica linguistica (fonetica, lessicale, grammaticale) da quelle senza. Le isoglosse IE possono essere di origine filogenetica (comune discendenza) o areale (diffusione per contatto).

Laringali: Serie di fonemi consonantici ricostruiti per il PIE da Ferdinand de Saussure (1879) e confermati dalla scoperta dell'ittita. Si ricostruiscono tipicamente come $*H_1$ (fricativa glottidale), $*H_2$ (fricativa faringale/uvulare) e $*H_3$ (fricativa labiovelare). Conservate come h- in ittita, perse altrove con effetti residui sulle vocali adiacenti.

Legge di Grimm: Insieme sistematico di cambiamenti consonantici che differenziano il proto-germanico dalle altre lingue IE. Formulata da Jacob Grimm nel 1822. Le occlusive sorde PIE > fricative sorde germ.; occlusive sonore PIE > occlusive sorde germ.; sonore aspirate PIE > fricative sonore germ.

Lineare B: Scrittura sillabica usata nei palazzi micenei (1400-1100 a.C.) per l'amministrazione. Decifrata da Michael Ventris nel 1952. Attesta il greco miceneo, la più antica forma di greco documentata, risalente di 500 anni rispetto a Omero.

MRCA: Most Recent Common Ancestor. L'antenato più recente condiviso da un insieme di individui o sequenze. Nella coalescenza di Kingman, il tempo all'MRCA di un campione di n individui è in media 4Ne (per N grande).

PIE (Proto-Indo-European): La lingua ancestrale ricostruita della famiglia indoeuropea. Non direttamente attestata (nessuna iscrizione PIE è mai stata trovata), ma ricostruita attraverso la comparazione sistematica di tutte le lingue IE attestate usando il metodo comparativo della linguistica storica.

qpAdm: Software (parte di ADMIXTOOLS, Patterson et al.) che stima le proporzioni di miscela di una popolazione "sorgente" modellata come combinazione lineare di popolazioni di riferimento. Usa le f_4-statistiche come base per la stima.

Sintashta: Cultura archeologica delle steppe meridionali (Urali meridionali, 2100-1800 a.C.). Prima cultura a produrre carri leggeri con raggi (2000-1900 a.C.). Genomicamente vicina allo Yamnaya (Steppe_MLBA), candidata principale come vettore del proto-indo-iranico.

Sprachbund: Unione linguistica. Insieme di lingue che condividono caratteristiche strutturali per contatto geografico prolungato, non per discendenza comune. La Sprachbund balcanica (greco, bulgaro, romeno, albanese) è l'esempio classico.

Trifunzione (Dumézil): L'ideologia tripartita proto-indoeuropea proposta da Georges Dumézil: funzione sovrana (sacerdozio/legge), funzione guerriera (forza militare), funzione produttiva (abbondanza/fertilità). Riscontrabile nelle tradizioni mitologiche, religiose e sociali di molteplici rami IE.

Urheimat: Termine tedesco per "patria originaria". Nel contesto delle origini IE, la regione geografica in cui si trovava la comunità di parlanti il PIE prima della dispersione. Il candidato principale delle ricerche attuali è la steppa pontico-caspia.

WHG (Western Hunter-Gatherers): Componente genomica dei cacciatori-raccoglitori dell'Europa occidentale del Mesolitico (Loschbour, Motala, La Braña). Si distinguono dagli EHG per maggiore affinità con le popolazioni dell'Europa meridionale e minore componente ANE.

Yamnaya: Cultura delle steppe (3300-2500 a.C.) distribuita dalla Ucraina meridionale al Kazakhstan occidentale, identificata da kurgan (tumuli funerari), sepolture ocra, pastoralismo mobile. Il principale candidato genomico e archeologico come vettore della dispersione IE in Europa.

Appendice O — Il problema della glottocronologia: limiti e revisioni

La glottocronologia — il metodo per datare la separazione delle lingue sulla base del tasso di sostituzione del lessico di base (lista di Swadesh) — è stata proposta da Morris Swadesh negli anni Cinquanta e ha avuto un periodo di grande influenza prima di essere ridimensionata dalla critica metodologica. Il principio fondamentale è che il lessico di base (pronomi, numerali, parti del corpo, termini fondamentali) si rinnova a un tasso costante, ipotizzato da Swadesh intorno al 14-19% per millennio. Analogamente all'orologio molecolare in biologia, la percentuale di lessico condiviso (cognati) tra due lingue permetterebbe di stimare il tempo di separazione.

La glottocronologia classica soffre di almeno quattro problemi metodologici seri. Primo, il tasso di sostituzione lessicale non è costante tra le lingue: alcune lingue

mostrano tassi molto più alti (lingue con intensa storia di contatto) o molto più bassi (lingue con tradizione letteraria conservativa). Secondo, il riconoscimento dei cognati è soggettivo e dipende dallo stato della ricostruzione comparativa, producendo circolarità parziale. Terzo, il prestito lessicale — particolarmente tra lingue adstrate — mima la cognanza e falsifica le stime. Quarto, il "tasso" di sostituzione varia sistematicamente con la frequenza d'uso delle parole (parole ad alta frequenza si sostituiscono più lentamente).

La filogenesi bayesiana moderna (BEAST, MrBayes) ha sostituito la glottocronologia classica con un approccio più sofisticato che modella esplicitamente la variazione del tasso, usa la presenza/assenza di cognati invece della semplice percentuale, e incorpora vincoli di calibrazione (attestazioni storiche datate) per ancorare l'orologio molecolare linguistico a date assolute. I risultati (Gray e Atkinson 2003, Chang et al. 2015, Heggarty et al. 2023) mostrano che la glottocronologia classica sottostimava sistematicamente le date di separazione dei rami IE, perché non teneva conto della variazione del tasso, e che le date bayesiane sono generalmente più antiche di 500-1500 anni rispetto alle stime glottocronologiche classiche.

Nodo IE	Glottocron. Swadesh	BEAST Chang 2015	BEAST Heggarty 2023	Attesta. storica (bound)
Radice PIE	~4000-3000 a.C.	4800-3000 a.C. (95% HPD)	8120-5940 a.C. (95% HPD)	Non attestato
Anatolico (separaz.)	~4000 a.C. (stimata)	5400-4200 a.C.	6700-5200 a.C.	Ittita ~1650 a.C.
Indo-iranico (radice)	~2500 a.C.	3200-2200 a.C.	3900-2800 a.C.	Vedico ~1200 a.C.
Greco (radice)	~2000 a.C.	2800-1900 a.C.	3500-2500 a.C.	Miceneo ~1400 a.C.
Proto-germanico	~500-0 a.C.	1800-500 a.C.	2400-1200 a.C.	Gotico ~350 d.C.
Proto-slavo	~500-0 a.C.	1500-500 a.C.	2200-1000 a.C.	Paleoslavo ~860 d.C.
Proto-celtico	~1000 a.C.	2200-900 a.C.	2800-1500 a.C.	Celtico continentale ~300 a.C.

Appendice P — Sintesi bibliografica critica: le opere fondamentali

La seguente rassegna bibliografica sistematica presenta le opere fondamentali nella storia delle ricerche sulle origini IE, organizzate per periodo e disciplina, con un giudizio critico sul contributo e sullo stato attuale.

Jones, Sir William (1786). "The Third Anniversary Discourse" alla Asiatic Society of Bengal. — La prima formulazione sistematica dell'ipotesi di parentela tra sanscrito, greco, latino, gotico, celtico e persiano. Non una pubblicazione scientifica nel senso moderno, ma il testo fondatore di tutta la linguistica comparativa IE. Il contributo fondamentale è la proposta del metodo: le affinità tra le lingue sono così sistematiche da postulare una fonte comune.

Schleicher, August (1861-1862). Compendium der vergleichenden Grammatik der indogermanischen Sprachen. — La prima sintesi sistematica della grammatica comparativa IE, con la proposta del Stammbaum-Theorie (teoria dell'albero). Schleicher tentò anche di scrivere una favola in PIE ricostruito ("Avis akvāsas ka"), il primo tentativo di usare il PIE come lingua produttiva.

Saussure, Ferdinand de (1879). Mémoire sur le système primitif des voyelles dans les langues indo-européennes. — La proposta delle laringali, formulata da Saussure ventiduenne. Uno dei contributi intellettuali più profondi e originali della linguistica IE. Confermata dalla scoperta dell'ittita quarantotto anni dopo.

Gimbutas, Marija (1970, 1973, 1977). Serie di articoli e libri che articolano la "Kurgan Hypothesis". — Il framework geografico che localizza l'Urheimat nelle steppe pontico-caspiche. Controversa al momento della formulazione per la tesi della "distruzione della civiltà della dea" (le culture neolitiche europee come culture matriarcali distrutte dall'invasione patriarcale delle steppe), accettata nella componente geografica e demografica dai dati aDNA più recenti.

Renfrew, Colin (1987). Archaeology and Language: The Puzzle of Indo-European Origins. — La proposta dell'ipotesi anatolica, che localizza l'Urheimat in Anatolia e identifica la neolitizzazione come vettore. Fortemente indebolita dai dati aDNA post-2015, rimane un testo fondamentale per la sua sistematizzazione del problema e per la sua critica del modello kurganico.

Gamkrelidze, Thomas e Ivanov, Vjačeslav (1984). Indoevropejskij jazyk i indoevropejcy (pubblicato in inglese 1995). — La proposta dell'Urheimat armeno-caucasico e della "glottalic theory" del sistema consonantico PIE. Oggi minoritaria nel suo insieme, ma ha contribuito alla discussione sulla tipologia fonologica del PIE e sulla localizzazione del ramo anatolico.

Dumézil, Georges (1958, 1968-73). L'idéologie tripartie des Indo-Européens; Mythe et épopée (3 voll.). — Il programma di ricostruzione mitologica e ideologica PIE attraverso la "trifunzione". Opera di straordinaria erudizione e sistematicità, influente sia in Indo-European Studies che in mitologia comparata e strutturalismo.

Anthony, David (2007). The Horse, the Wheel, and Language. — La sintesi più accessibile e aggiornata dell'evidenza archeologica per l'ipotesi delle steppe. Integra archeologia, linguistica e (precocemente per il 2007) alcune considerazioni genomiche. Considerato il testo di riferimento per i non-specialisti sull'ipotesi kurganica.

Haak, Wolfgang et al. (2015). "Massive migration from the steppe was a source for Indo-European languages in Europe." Nature. — Il paper fondamentale che ha introdotto la componente "ancestralità steppa" e ha dimostrato l'espansione demografica dall'est con dati aDNA di 69 individui europei antichi. Ha trasformato il dibattito.

Allentoft, Morten et al. (2015). "Population genomics of Bronze Age Eurasia." Nature. — Contemporaneo a Haak et al., analizza 101 individui del Bronze Age eurasiatico e conferma l'espansione steppa e l'identificazione Sintashta-proto-Indo-iranico.

Narasimhan, Vagheesh et al. (2019). "The formation of human populations in South and Central Asia." Science. — L'estensione dell'analisi aDNA all'Asia centrale e al subcontinente indiano. Documenta l'assenza di componente steppa nelle popolazioni harappane e la sua comparsa successiva, supportando il modello Sintashta → India per l'indo-ariano.

Heggarty, Paul et al. (2023). "Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages." Science. — L'analisi filogenetica bayesiana più aggiornata, con 161 lingue e 4500 caratteri lessicali.

Produce date più antiche del previsto (radice ~7100-5900 a.C.) e struttura dell'albero ibrida tra Anatolia e steppe.

Zhang, Fan et al. (2021). "The genomic origins of the Bronze Age Tarim Basin mummies." Nature. — L'analisi aDNA delle mummie del Tarim che rivela un profilo ANE quasi puro, complicando il quadro delle origini del tocarico.

Appendice Q — Tavola cronologica sinottica delle origini IE

Data (a.C.)	Evento archeologico	Evento linguistico (ipotetico)	Evento genomico	Ramo IE coinvolto
~5000-4700	Neolitico steppe, prime domesticaz. bovini	Pre-PIE nelle steppe pontiche	EHG puro, CHG assente	Nessuno (pre-IE)
~4700-4000	Cultura Khvalynsk (medio Volga)	PIE arcaico in formazione	EHG 80% + CHG 15%	Nessuno (proto-IE)
~4500-3700	Cultura Sredny Stog (Dnepr-Don)	PIE arcaico in espansione	EHG 68% + CHG 30%	Nessuno (proto-IE)
~3700-3000	Cultura Maykop (Caucaso N.)	Contatto IE/non-IE nel Caucaso	CHG 60%+ (non-IE prob.)	Fonte CHG nello Yamnaya
~3300-2700	Cultura Yamnaya (steppe)	PIE maturo	EHG 52% + CHG 43%	Proto-IE pre-dispersione
~3300-2500	Cultura Afanasievo (Altai)	Proto-tocarico verso est	Yamnaya-like (EHG 62%)	Tocarico
~3100-2500	Yamnaya si espande verso ovest	Proto-europeo IE	Steppa in Europa centr.	Proto-CW (futuro N.Europ.)
~2900-2300	Cultura Corded Ware (Europa C.)	Proto-IE europeo differenziato	Steppa 40-75%	Futuro germ., BS, celt., ital.
~2500-1800	Cultura Bell Beaker (Occidente)	Proto-celtico verso atlantico	Steppa 25-50%	Celtico, forse proto-ital.
~2100-1800	Sintashta (Urali S.)	Proto-Indo-iranico cristallizza	Steppa_MLBA R1a-Z93	Indo-iranico
~2100-1800	Sintashta: primo carro a raggi	PIE *roteh$_2$ (ruota a raggi)	Vedi sopra	Indo-iranico (carro)
~1800-1400	Andronovo (Kazakhstan)	Proto-indo-iranico si espande	R1a-Z93 dominante	Indo-iranico (espansione)

Data (a.C.)	Evento archeologico	Evento linguistico (ipotetico)	Evento genomico	Ramo IE coinvolto
~1650	Ittita: prime tavolette	Primo testo IE attestato	Anatolia: poca steppa	Anatolico (attestazione)
~1350	Micene: tavolette Lineare B	Proto-greco attestato	Egeo: 10-15% steppa	Greco (attestazione)
~1200	Rigveda (composizione)	Vedico arcaico attestato	India N.: steppa in arrivo	Indo-ariano (attestazione)
~1200-700	Collasso del Bronze Age	Discontinuità linguistiche	Riorganizzazioni genomiche	Vari
~1000-500	Età del Ferro europea	Differenz. completa rami IE	Profili moderni emergono	Tutti i rami

CAPITOLO XVI — IL RAMO BALTO-SLAVO: IL PIU' CONSERVATIVO DELLA FAMIGLIA INDOEUROPEA

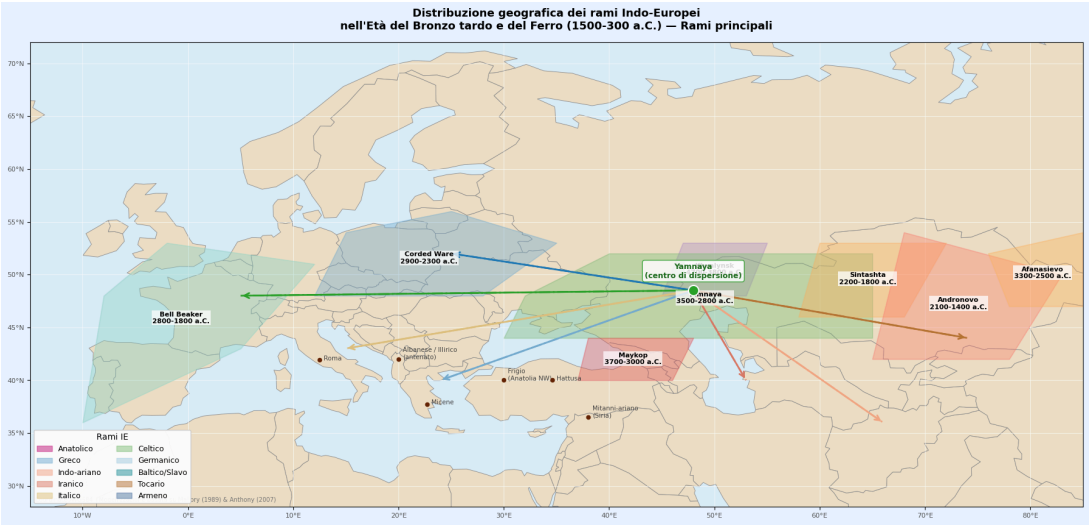


Figura C — Mappa delle dispersioni dei rami indoeuropei con frecce per i singoli rami. Colori distinti indicano: indo-iranico (verde), greco-armeno-albanese (blu), italico-celtico (arancio), germanico (giallo), balto-slavo (rosso), tocarico (viola), anatolico (marrone). Elaborazione originale.

XVI.1 Struttura interna e posizione filogenetica del Balto-Slavo

XVI.1.1 Il dibattito sull'unita' balto-slava: argomenti pro e contro

Il ramo balto-slavo dell'albero indoeuropeo comprende due sotto-rami principali: le lingue baltiche (lituano, lettone, e il prussiano antico, estinto nel XVII secolo) e le lingue slave (russo, polacco, ceco, slovacco, serbo-croato, bulgaro, sloveno, ucraino, bielorusso, macedone, e numerose lingue minori). La questione se Baltico e Slavo costituiscano un nodo filogenetico reale — un sotto-ramo dell'IE che ha un antenato comune esclusivo distinto da tutte le altre lingue IE — e' stata dibattuta per circa due secoli senza risoluzione unanime. I sostenitori dell'unita' balto-slava (tra cui August Schleicher, Karl Brugmann e piu' recentemente Frederick Kortlandt) adducono un corpus di isoglosse morfologiche, fonologiche e lessicali condivise che, a loro avviso, non possono essere spiegate se non come innovazioni condivise del Proto-Balto-Slavo (PBS). Gli oppositori (tra cui Antoine Meillet e in parte Antoine Ernout) interpretano le stesse isoglosse come fenomeni di convergenza areale tra due rami originariamente indipendenti che si sono sviluppati in adiacenza geografica per millenni.

Le isoglosse principali che unificano Baltico e Slavo in un nodo PBS sono di quattro tipi. (1) La monoftonghizzazione dei dittonghi: il PBS ha sistematicamente monoftonghizzato i dittonghi PIE *ei e *ai in modo parallelo nelle due sub-famiglie, un cambiamento fonologico specifico non condiviso con nessun altro ramo IE nella stessa forma. (2) Il sistema dei tempi verbali: sia il Baltico che lo Slavo hanno innovato in modo parallelo il sistema verbale PIE, eliminando il sistema quadripartito di aoristo, perfetto, imperfetto e presente in favore di un sistema basato sull'aspetto (perfettivo/imperfettivo) piuttosto che sul tempo assoluto — un'innovazione morfologica molto specifica che non ha paralleli diretti in nessun altro ramo. (3) La palatalizzazione delle velari: il PBS appartiene al gruppo satem (assieme all'indo-iranico e all'armeno), con le velari PIE *k e *g palatalizzate in spiranti o sibilanti — tuttavia questa convergenza con l'indo-iranico e' interpretata come innovazione areale indipendente e non come innovazione PBS specifica. (4) La desinenza del genitivo singolare dei temi in -o-: sia il Baltico (lituano -o) che lo Slavo (antico slavo ecclesiastico -a) derivano da una desinenza comune PBS che differisce da quelle del greco, del latino e del germanico.

L'analisi bayesiana della filogenesi IE (Chang et al. 2015; Heggarty et al. 2023) assegna alta probabilita' posteriore (> 0.85) al nodo balto-slavo, suggerendo che l'ipotesi dell'unita' filogenetica sia preferita dalla struttura cladistica del lessico. Tuttavia, la lunghezza del ramo interno che porta al nodo PBS e' significativamente inferiore a quella di altri nodi IE ben consolidati (come il nodo indo-iranico), compatibilmente con la possibilita' che il PBS sia un nodo reale ma con un'esistenza come entita' linguistica relativamente breve — un'ipotesi detta dello "short branch scenario" PBS.

Isoglossa	Baltico	Slavo	Parallelo in altri rami IE	Interpretazione prevalente
Monoftongazione *ei > i:	lit. iemas < *ghei-mo-	sl. zima "inverno"	Nessun parallelo esatto	Innovazione PBS condivisa
Sistema aspettuale verbo	Distinzione pf./impf.	Distinzione pf./impf. sistematica	Assente come sistema	Innovazione PBS condivisa
Gen. sg. temi -o-	Lit. -o, lett. -a	ACS -a, russo -a	Gr. -ou, lat. -i, got. -is	Innovazione PBS condivisa

Isoglossa	Baltico	Slavo	Parallelo in altri rami IE	Interpretazione prevalente
Palatalizzazione velari	s < *k (gruppo satem)	s, z < *k, *g	Indo-ir., armeno (satem)	Convergenza areale satem
Riflessivo *swe-	Lit. save, lett. sevi	Sl. sebe, russo sebja	Lat. se, gr. (he-)	Probabile PBS condivisa
Participio presente	Desinenza -nt- attiva	Desinenza -nt- attiva	Largamente condiviso IE	Non diagnostica per PBS
Suffisso diminutivo *-ik-	Produttivo in Baltico	Molto produttivo in Slavo	Raro altrove	Possibile innovazione PBS

XVI.1.2 Il Proto-Balto-Slavo: ricostruzione del sistema fonologico

La ricostruzione del Proto-Balto-Slavo (PBS) e' condizionata dall'ambiguita' relativa al suo status di nodo reale o costrutto areale. Adottando l'ipotesi del nodo PBS come operativamente valida, il suo sistema fonologico si differenziava dal PIE tardo per le seguenti innovazioni principali. (1) La regola di Winter: le vocali brevi PIE si allungavano davanti alle occlusive sorde non aspirate *p, *t, *k quando precedevano le stesse occlusive. Questa regola, formulata da Werner Winter nel 1978 e poi affinata da Kortlandt e Rasmussen, produce alternanze di quantita' vocalica che distinguono il PBS da tutti gli altri rami IE e che lasciano tracce sia in Baltico (toni distintivi nel lituano moderno) che in Slavo (vocali "lunghe" nel proto-slavo). (2) Il trattamento dei gruppi consonantici PIE *tl e *dl: il PBS li riduce a *l (a differenza del germanico che li conserva come *tl > dl), come mostrato dal lituano atzala "ramo" vs tedesco Zweig e russo zolo- in composti. (3) Il sistema dei laringali: il PBS risolve i laringali PIE *H1, *H2, *H3 davanti a consonante producendo vocali brevi di diversa qualita' (-e-, -a-, -o- rispettivamente), con tracce conservate nell'alternanza vocalica che permea la morfologia baltica e slava.

Il sistema accentuale del PBS e' di importanza capitale per la comprensione della fonologia storica del ramo. Il lituano moderno e il lettone conservano una distinzione fonologica di tono (acuto vs. circonflesso) che deriva dall'accentuazione del PBS e, attraverso di essa, dall'accentuazione del PIE. Ferdinand de Saussure (1896) e Karl Brugmann avevano gia' riconosciuto che la distinzione tonica

balto-slava corrispondeva sistematicamente alle distinzioni di quantita' vocalica del greco classico — un'osservazione che ha portato alla formulazione della "legge di de Saussure" nella sua applicazione balto-slava. La corrispondenza piu' significativa e' quella tra il tono acuto baltico (intonazione "rising") e il greco con vocale lunga, e tra il tono circonflesso baltico e il greco con dittongo o vocale di diversa origine. Questa corrispondenza e' interpretata come conservazione congiunta di una distinzione presente nel PIE tardivo, forse riflettente la distinzione tra laringali *H1 e *H2/*H3 in posizione post-vocalica.

Fenomeno PBS	Esempio Lituano	Esempio Slavo	PIE ricostruito	Commento
Regola di Winter: VT > V:T	lit. duoti "dare" (d lungo)	sl. dati (d lungo)	*deh3-	Allungamento pre-consonantico
Laringale H1 > e	lit. edmi "mangio"	sl. jasti "mangiare"	*H1ed-mi	H1 > zero con colorazione e
Laringale H2 > a	lit. akis "occhio"	sl. oko "occhio"	*H2okw-i-	H2 > zero con colorazione a
Laringale H3 > o	lit. olas "grotta"	sl. ulo "arnia"	*H3ul-	H3 > zero con colorazione o
Tono acuto vs. circonflesso	lit. rankos (genit.) acuto	proto-sl. *raka (nom.)	*reh2k-a-	Conservazione distinzione PIE
*tl, *dl > l	lit. atzala < *at-tla-	russo -l- in participi	*-tlo-	Semplificazione consonantica PBS
*sr > str	lit. srove > srove	sl. struja "corrente"	*srou-	Metatesi consonantica PBS

XVI.2 Le lingue baltiche: lituano, lettone e il prussiano antico

XVI.2.1 Il lituano: la lingua IE piu' arcaica tra le lingue vive

Il lituano occupa una posizione del tutto eccezionale tra le lingue indoeuropee vive per il grado straordinario di arcaicita' che conserva sia nella fonologia che nella morfologia. August Schleicher, il fondatore della linguistica comparata sistematica, identifico' nel lituano il miglior testimone vivente del PIE, e questa valutazione — pur con le necessarie qualificazioni metodologiche — rimane sostanzialmente valida. La morfologia nominale del lituano conserva sette casi (nominativo, genitivo,

dativo, accusativo, strumentale, locativo e vocativo) contro i cinque del latino (perdita di strumentale e locativo) e i due del neogreco moderno, avvicinandosi al sistema degli otto casi ricostruito per il PIE (nominativo, genitivo, dativo, accusativo, strumentale, locativo, ablativo e vocativo — con il lituano che ha fuso ablativo e genitivo). Il sistema verbale del lituano conserva distinzioni aspettuali, di diatesi media e di modo che riflettono lo stadio PIE con un grado di fedeltia' superiore a quello del latino arcaico o del greco omerico.

La posizione del lituano nella ricostruzione del PIE e' metodologicamente fondamentale per un aspetto specifico: il sistema dei toni. Come gia' discusso in XVI.1.2, il lituano conserva una distinzione fonologica di intonazione (tono acuto vs. circonflesso) sulle sillabe lunga e sui dittonghi, distinzione che, attraverso la corrispondenza di de Saussure, permette di ricostruire aspetti dell'accentuazione PIE che sarebbero altrimenti inaccessibili. La forma verbale lituano yra "e'" (3a persona singolare del presente di essere) e' spesso citata come esempio paradigmatico di arcaicita': e' derivabile direttamente da *H1es-ti (PIE per "e'") attraverso cambiamenti fonologici regolari che includono la perdita della laringale con allungamento compensatorio, la perdita della desinenza -ti in posizione finale assoluta, e l'evoluzione della vocale di radice. La corrispondenza e' quasi biunivoca: ogni fono del lituano moderno puo' essere ricondotto a un fono PIE ricostruito attraverso regole fonetiche sistematiche.

Il lituano e' parlato da circa 3 milioni di persone, principalmente in Lituania (dove e' lingua ufficiale) e da comunita' diasporiche in Polonia, Russia, e numerosi paesi occidentali. La standardizzazione della lingua letteraria risale al XVI secolo con la traduzione della Bibbia e dei catechismi luterani, e la continuita' ininterrotta del patrimonio letterario ha contribuito alla conservazione di forme arcaiche che in molte altre lingue IE sono andate perse. Il primo documento scritto in lituano e' il Catechismo di Martynas Mazyvdas (1547), in un'ortografia latinizzata che conserva distinzioni fonologiche non presenti nel latino medievale.

Forma lituana	Significato	PIE ricostruito	Latino corrisp.	Greco corrisp.	Commento
yra	e' (3sg)	*H1esti	est	esti	Massima arcaicita'
esmi	sono (1sg)	*H1esmi	sum (< *esom)	eimi	Desinenza -mi arcaica

Forma lituana	Significato	PIE ricostruito	Latino corrisp.	Greco corrisp.	Commento
rankas (acc. pl.)	mani	*ronkans	manus	kheiras	Acc. pl. in -ns
sunus (nom. sg.)	figlio	*sunus	filius	hyios	Tema in -u conservato
devyni	nove	*H1newn	novem	ennea	Vocale iniziale IE *e
vilkas	lupo	*wlkwos	lupus (< *wlkwos)	lykos	Conserva -lk-
naujas	nuovo	*newos	novus	neos	Dittongo au < *ew
diena	giorno	*dien-	dies	Zeus (gen. Dios)	Radice *dyew-

XVI.2.2 Il lettone e la sua posizione rispetto al lituano

Il lettone — parlato da circa 1.5 milioni di persone in Lettonia e in diaspora — e' la seconda lingua baltica viva e la lingua piu' strettamente imparentata al lituano tra quelle con un corpus letterario sostanziale. Nonostante la vicinanza genetica, il lettone ha subito innovazioni fonologiche significative che lo differenziano notevolmente dal lituano, rendendo le due lingue molto piu' distanti di quanto la genetica comune suggerirebbe. Le innovazioni piu' rilevanti del lettone rispetto al lituano includono: (1) la perdita del tono acuto distinto, con fusione dei due toni lituani (acuto e circonflesso) in una distinzione tripartita in lettone (staccato, esteso, spezzato) che riflette in parte una ricategorizzazione fonologica; (2) la perdita dell'intonazione su sillabe aperte non accentate, con caduta di vocali finali brevi che ha prodotto un sistema morfologico piu' eroso del lituano; (3) l'evoluzione del sistema degli infiniti e dei participi in direzione di una maggiore analiticita' rispetto al lituano; (4) influenze del substrato finno-ugrico (livone, finlandese meridionale) particolarmente evidenti nel vocabolario relativo all'ambiente naturale e alla pesca.

Il confronto lettone-lituano e' metodologicamente prezioso per la ricostruzione del Proto-Baltico perche' le due lingue, essendo fratelle, permettono di triangolare le forme del loro antenato comune con maggiore precisione di quanto sia possibile con una sola lingua. Molte isoglosse conservate in lituano ma perdute in lettone (o viceversa) possono essere ricondotte a innovazioni baltiche settentrionali

(prelettoniche) o meridionali (prelituane), arricchendo la nostra comprensione della struttura interna del Baltico. Il prussiano antico — la terza lingua baltica, estinta alla fine del XVII secolo per sostituzione con il tedesco dopo la prussianizzazione — conserva alcune forme arcaiche perdute sia in lituano che in lettone, e il suo corpus (principalmente il Catechismo di Enchiridion del 1561) e' stato intensamente studiato per la ricostruzione del Proto-Baltico.

XVI.3 Le lingue slave: dalla dispersione del Proto-Slavo alla molteplicita' contemporanea

XVI.3.1 Il Proto-Slavo e la sua datazione: evidence linguistica e archeogenomica

Il Proto-Slavo — antenato comune di tutte le lingue slave — e' generalmente datato tra il III e il VI secolo d.C., con la sua dispersione geografica principale tra il VI e il IX secolo d.C. Questa datazione relativamente tarda (rispetto ad altri rami IE) e' supportata da tre linee di evidenza convergenti. (1) La bassa diversita' interna delle lingue slave: tutte le lingue slave moderne sono mutuamente intelligibili (con variabile difficolta') e condividono un grado di somiglianza lessicale molto superiore a quello delle lingue germaniche, delle lingue romanze o delle lingue iraniche — compatibile con una diversificazione iniziata circa 1500 anni fa. (2) Le attestazioni storiche: la prima lingua slava documentata e' il veccchio slavo ecclesiastico (Old Church Slavonic, OCS), utilizzato da Cirillo e Metodio nella traduzione della Bibbia per i Moravi intorno all'863 d.C. — una lingua che riflette lo slavo parlato in Macedonia meridionale nel IX secolo. (3) I dati di filogeografia molecolare: studi genomici sulle popolazioni slave (Lazaridis et al. 2016; Freilich et al. 2021) identificano una componente genetica condivisa tra gli slavi moderni che, calibrata con dati archeogenomici, e' compatibile con una espansione rapida da un'area sorgente nell'Europa orientale a partire dal V-VI secolo d.C.

La localizzazione dell'Urheimat del Proto-Slavo — la sua patria di origine prima della dispersione — e' dibattuta tra un'area a nord dei Carpazi (odierna Polonia meridionale, Ucraina occidentale e Bielorussia meridionale), proposta da Trubeckoj e sostenuta da molti slavisti, e un'area a sud dei Carpazi (Pannonia, odierna Ungheria), proposta da pochi ma non priva di supporto lessicale. La maggioranza degli studiosi moderni favorisce la localizzazione nord-carpatica per due ragioni

principali: (1) il lessico del Proto-Slavo ricostruito include numerosi termini per la flora e la fauna dell'Europa orientale temperata (quercia, faggio, lupo, castoro, cavallo) ma e' privo di termini sicuri per il Mediterraneo; (2) le prime menzioni di Slavi nelle fonti greche e latine (Procopio, Giordane, VI secolo d.C.) li localizzano nell'area tra il Vistola e il Dnepr, compatibile con una migrazione da nord-est verso la penisola balcanica e verso le pianure pannoniche nel periodo delle grandi migrazioni (IV-VII secolo d.C.).

Gruppo slavo	Lingue principali	Area geografica	Parlanti (milioni)	Prima attestazione
Slavo orientale	Russo, Ucraino, Bielorusso	Europa orientale, Russia	210	Cronaca di Nestore, XI sec.
Slavo occidentale	Polacco, Ceco, Slovacco, Sorabo	Europa centrale	57	Boemia, X-XI sec.
Slavo meridionale	Serbo-croato, Bulgaro, Sloveno, Macedone	Balcani, Slovenia	35	OCS, IX sec. (Cirillo-Metodio)
Lusaziano (Sorabo)	Alto e basso sorabo	Germania orientale	0.07	XVI sec. (residuale)
OCS (estinto)	Vecchio Slavo Eccl.	Macedonia/Bulgaria	Estinto	863 d.C. (Vangelo)
Casciubico	Casciubico	Pomerania polacca	0.1	XIX sec. (letterario)

XVI.3.2 La morfologia del vecchio slavo ecclesiastico: arcaismi PIE e innovazioni slave

Il vecchio slavo ecclesiastico (OCS) rappresenta il documento piu' arcaico del ramo slavo e la fonte primaria per la ricostruzione del Proto-Slavo. La morfologia nominale dell'OCS conserva sette casi (nominativo, accusativo, genitivo, dativo, strumentale, locativo e vocativo), perdendo solo l'ablativo PIE (fuso con il genitivo come in lituano) e conservando una distinzione duale (forme speciali per i referenti in coppia: occhi, mani, orecchie) che e' arcaismo PIE preservato anche in greco e in sanscrito ma perduto dal latino. Il sistema verbale dell'OCS include un aoristo sigmativo (perfettamente parallelo all'aoristo greco) che riflette la categoria PIE dell'aoristo, una distinzione perfettivo/imperfettivo che costituisce la principale

innovazione del PBS, e un participio presente attivo con desinenza -nt- che risale direttamente al PIE *-nt- (lat. -ns/-ntis, gr. -nt-).

Il lessico dell'OCS e' ricco di prestiti da lingue con cui i Proto-Slavi furono in contatto durante e dopo la loro espansione. I prestiti germanici riflettono contatti con le tribu' germaniche dell'Europa orientale (goti, vandali) nel periodo delle grandi migrazioni: OCS kupiti "comprare" (dal gotico kaufen), OCS hlebu "pane" (dal gotico hlaifs), OCS chlamu "vergogna" (dal gotico hlamos). I prestiti greci riflettono l'influenza della chiesa ortodossa e dell'Impero Bizantino: OCS angelu "angelo", OCS evangelijie "vangelo", OCS patriarchu "patriarca". I prestiti iranici riflettono contatti prolungati con popoli delle steppe di lingua iranica (Sciti, Sarmati, Alani): OCS bugu "dio" (cf. avestico baga- "signore"), OCS rajyu "paradiso" (cf. avestico rayu "ricchezza"), OCS sanu "sogno" (cf. ossetico san).

La prosopologia del PBS ricostruito, con le sue innovazioni fonologiche sistematiche (le tre palatalizzazioni slave, la metatesi dei liquidi, la regola delle sillabe aperte), costituisce uno dei sistemi di corrispondenze fonetiche piu' elegantemente documentati nella linguistica storica. La "legge delle sillabe aperte" del Proto-Slavo — secondo cui ogni sillaba del PS terminava in vocale (erano proibite le sillabe chiuse) — ha prodotto una serie di cambiamenti sistematici che includono la metatesi dei gruppi liquida-vocale (*TorT > *ToroT, dove T = consonante), l'epentesi di vocali epentetiche davanti a gruppi consonantici, e la perdita delle occlusive finali. Questo principio fonologico, ricostruito da Jakobson e sviluppato da Shevelov, spiega in modo unificato una grande quantita' di dati slave che sarebbero altrimenti difficilmente sistematizzabili.

Processo fonologico PBS	PIE/PBS ricostruito	OCS risultante	Russo moderno	Commento
1a palatalizzazione k > ch	*kena "donna"	zena	zhena	Davanti a e/i/j orig.
2a palatalizzazione k > ts	*keina > *keina	tsena "prezzo"	tsena	Davanti a e/i di orig. dittongale
Metatesi liquide *TorT	*gord- "recinto"	gradu "citta'"	gorod	Metatesi balcanica
Metatesi liquide *TolT	*bolto "palude"	blato	boloto (form. arcaica)	Metatesi balcanica
Monoftong. *ai > e	*ghaino "femm."	zena "donna"	zhena	Parallelo in Baltico

Processo fonologico PBS	PIE/PBS ricostruito	OCS risultante	Russo moderno	Commento
Monoftong. *au > u	*lausos "vuoto"	lyudu "popolo"	ljud	Monoftong. PBS
Regola sillabe aperte	*vert- "girare"	vruteti (epentesi)	vertet'	Proibiz. sill. chiuse
Nasalizzazione *en, *on	*zengti "portare"	nesti (cons. rif. nasale)	nesti	Nasali PBS poi denasalizzate

XVI.4 La genomica delle popolazioni baltiche e slave: storia demografica e aDNA

La storia demografica delle popolazioni che parlano lingue balto-slave e' ricostruibile attraverso una combinazione di dati aDNA, di filogeografia di aplogruppi del cromosoma Y e del mtDNA, e di inferenza bayesiana sulla dimensione efficace di popolazione. Le popolazioni baltiche moderne (lituani, lettoni) mostrano nei loro profili genomici una composizione che riflette la stratificazione demografica dell'Europa nordorientale: un substrato di cacciatori-raccoglitori baltici (WHG e EHG) — particolarmente pronunciato rispetto alle popolazioni dell'Europa centrale —, una componente di agricoltori neolitici ridotta rispetto alla media europea (compatibile con una neolitizzazione tardiva e parziale), e una componente di ancestralita' delle steppe (steppe ancestry) acquisita attraverso le espansioni della cultura Corded Ware. La componente EHG nelle popolazioni baltiche moderne e' notevolmente superiore alla media europea (circa 25-30% vs. 10-15%), coerente con la conservazione di una componente di cacciatori-raccoglitori baltici che non e' stata completamente diluita dalla neolitizzazione.

I dati aDNA del Neolitico e dell'Eta' del Bronzo delle regioni baltiche (pubblicati da Mittnik et al. 2018; Saag et al. 2019; Olalde et al. 2019) mostrano una discontinuita' genetica marcata tra gli agricoltori neolitici dell'area (con alta componente EEF, simili agli agricoltori dell'Europa centrale) e le popolazioni dell'Eta' del Bronzo (con alta componente steppa, introdotta attraverso la cultura Corded Ware intorno al 2800-2400 a.C.). Questa discontinuita' e' compatibile con l'arrivo del Proto-Balto-Slavo nell'area attraverso espansioni delle popolazioni Corded Ware. La ritenzione di una componente EHG particolarmente alta nelle popolazioni baltiche

moderne (rispetto a quelle germaniche o slave occidentali) suggerisce che la sostituzione demografica da parte della componente steppa non sia stata completa nell'area baltica come lo e' stata nell'Europa centrale.

Le popolazioni slave orientali (russi, ucraini, bielorussi) mostrano profili genomici con componenti che riflettono sia il substrato EHG delle steppe europee orientali — presente anche in popolazioni finno-ugrice con cui i Proto-Slavi orientali furono in contatto prolungato — sia una componente di steppe ancestry Corded Ware, e sia infine una componente di neolitico europeo centrale che riflette la diffusione degli agricoltori in Europa orientale. I dati di Lazaridis et al. (2016) e di Freilich et al. (2021) mostrano che la struttura genomica degli slavi orientali moderni e' piu' simile alle popolazioni dell'Europa orientale dell'Eta' del Bronzo finale che non alle popolazioni dell'Eta' del Ferro, suggerendo una sostanziale continuita' demografica nelle pianure est-europee per gli ultimi 3000 anni, con una sovraimpressione delle espansioni slave medievali sui substrati preesistenti.

Popolazione	EHG %	Steppa %	EEF %	WHG %	Altra %	Commento
Lituani moderni	28	38	25	7	2	Alta comp. EHG
Lettoni moderni	25	36	28	9	2	Simile lituano
Polacchi moderni	18	32	40	8	2	Piu' EEF
Russi moderni	22	30	35	5	8	8% comp. siberiana/ugrica
Ucraini moderni	20	33	38	7	2	Intermedio
Bulgari moderni	14	28	48	6	4	Alta comp. anatolica
Serbi moderni	15	27	47	7	4	Simile bulgari
Eneolitico Baltico	5	0	90	5	0	Pre-Corded Ware
Corded Ware Baltico	35	55	8	2	0	Arrivo steppa Corded Ware

CAPITOLO XVII — LA RICOSTRUZIONE SINTATTICA E PROSODICA DEL PROTO-INDOEUROPEO

XVII.1 La sintassi del PIE: ordine delle parole, particelle e prosodia frasale

XVII.1.1 L'ordine delle parole nel PIE: SOV o SVO? Il dibattito Lehmann-Watkins

La ricostruzione dell'ordine delle parole nel Proto-Indoeuropeo e' uno dei problemi piu' dibattuti della linguistica IE, con implicazioni metodologiche che travalicano la questione specifica e investono i fondamenti della ricostruzione sintattica comparata. Il dibattito e' stato avviato sistematicamente da Winfred Lehmann (1974) con la monografia "Proto-Indo-European Syntax", in cui Lehmann sosteneva che il PIE fosse una lingua SOV (Soggetto-Oggetto-Verbo) — dell'ordine cioe' del turco, del giapponese o del sanscrito arcaico — sulla base di tre argomenti: (1) le lingue IE piu' arcaiche (sanscrito vedico, ittita, latino) mostrano una forte preferenza per l'ordine SOV nei contesti narrativi non marcati; (2) la tipologia del PIE — con postposizioni, genitivo preverbale, ordine aggettivo-nome — e' pienamente coerente con una lingua SOV nella matrice di Greenberg; (3) la derivazione storica da SOV di tutti gli ordini osservati nelle lingue IE moderne richiede meno passi evolutivi della derivazione inversa.

La risposta di Watkins (1976) e di Szemerényi (1980) ha introdotto complessita' cruciali nel dibattito. Watkins ha mostrato che la cosiddetta "Legge di Wackernagel" — la collocazione delle particelle clitiche (enclitici) in seconda posizione nella frase (posizione Wackernagel) — impone vincoli all'ordine delle parole nel PIE che sono compatibili sia con SOV che con SVO, rendendo difficile estrarre un segnale sintattico pulito dall'ordine dei clitici. Szemerényi ha enfatizzato che i testi vedici e ittiti piu' antichi mostrano ordini liberi con una preferenza SOV nei contesti formali e rituali, ma SVO e VSO nei contesti narrativi vivaci — coerentemente con una distinzione tra ordine di default e ordine pragmatico. La conseguenza di questi studi e' che il PIE potrebbe essere stato una lingua a ordine libero con un default SOV, piuttosto che una lingua rigidamente SOV o SVO — un tipo non raro nelle lingue

del mondo (greco classico, sanscrito, latino sono esempi di lingue con "ordine libero pragmaticamente condizionato").

Il riesame bayesiano della questione da parte di Dunn et al. (2011) e di Gell-Mann et al. (2009) ha applicato metodi di filogenesi statistica agli ordini delle parole nelle lingue IE, ricostruendo lo stato ancestrale del nodo PIE con metodi parsimoniosi e bayesiani. Il risultato e' che il PIE ricostruito mostra alta probabilita' posteriore per l'ordine SOV (circa 0.65) ma non esclude SVO (circa 0.25), con il 10% di incertezza strutturale residua. Questo risultato e' interpretato come evidenza per un PIE prevalentemente SOV con variabilita' pragmatica — compatibile con la posizione di compromesso tra Lehmann e Watkins, e coerente con la tipologia osservata nelle lingue altamente flessive.

Lingua IE	Ordine prevalente	Ordine marcato	Flessione (gradi)	Fonte
Sanscrito vedico	SOV (narrativo)	SVO (dialogico)	Alta (8 casi)	Rigveda ~1200 a.C.
Ittita	SOV (formale)	SVO (quotidiano)	Media (6 casi)	Archivi hattusa ~1600 a.C.
Greco omerico	SVO/SOV libero	VSO (enfatico)	Alta (5 casi greci)	Iliade ~800 a.C.
Latino classico	SOV (formale)	SVO (prosa)	Alta (6 casi + voc.)	Cicerone, I sec. a.C.
Avestico	SOV	VSO (liturgico)	Alta (7 casi)	Gathas ~1000 a.C.
OCS	SVO/SOV libero	VSO (emfatico)	Media-alta (7 casi)	Testi IX sec.
Lituano moderno	SVO libero	SOV (letterario)	Alta (7 casi)	Standard moderno
Ricostruito PIE	SOV (prob. 0.65)	SVO (prob. 0.25)	Molto alta (8 casi)	Bayesian consensus

XVII.1.2 La legge di Wackernagel: clitici, seconda posizione e prosodia frasale

La legge di Wackernagel — formulata da Jacob Wackernagel nel 1892 e intitolata alla sua memoria — stabilisce che i morfemi clitici (particelle pronominali, avverbi di focalizzazione, connettivi) si collocano sistematicamente in seconda posizione nella frase (dopo il primo costituente fonologico), in una vasta gamma di lingue IE

arcaiche. La legge e' empiricamente documentata in sanscrito vedico, in greco omerico, in latino arcaico, in antico iranico, in ittita e in OCS, e la sua universalita' nel ramo IE arcaico suggerisce fortemente che essa rifletta una proprieta' del PIE stesso. La posizione Wackernagel e' "prosodica" nel senso che e' definita dal primo costituente fonologico della frase (che puo' essere una parola, un gruppo nominale, o anche un sintagma piu' lungo), non dall'ordine sintattico dei costituenti maggiori.

La formalizzazione moderna della legge di Wackernagel e' avvenuta nell'ambito della teoria della prosodia frasale (prosodic phonology) e della teoria dell'ottimalita'. Il clitico in posizione Wackernagel e' "debole" prosodicamente — non porta accento, si appoggia foneticamente alla parola che lo precede (enclisi) o a quella che lo segue (proclisi) — e occupa la seconda posizione per un'interazione tra vincoli prosodici (il clitico deve essere incorporato in un piede metrico) e vincoli sintattici (il clitico deve seguire il primo costituente della frase). La formalizzazione in termini di Teoria dell'Ottimalita' (Optimality Theory, Prince e Smolensky 1993) permette di derivare la collocazione Wackernagel come risultato ottimale di una gerarchia di vincoli universali con ranking specifico del PIE.

Un aspetto particolarmente rilevante della posizione Wackernagel e' il suo ruolo diagnostico per la ricostruzione dell'accentuazione PIE. I clitici in posizione Wackernagel sono per definizione privi di accento, e le particelle enclitiche sanscrite e greche che occupano questa posizione mostrano variazioni di qualita' vocalica (ablaut) che riflettono direttamente la posizione non accentata nel sistema apofonico PIE. Lo studio di Probert (2006) ha dimostrato che le particelle pronominali greche in posizione Wackernagel (enclitica me, se, oi, ni) corrispondono sistematicamente alle forme "ridotte" (grado zero) delle stesse radici nelle posizioni accentate, confermando che la posizione Wackernagel e' il contesto per eccellenza del grado zero apofonico nel PIE.

XVII.2 Il sistema prosodico del PIE: accento, quantita' e tono

XVII.2.1 L'accento PIE: natura, posizione e sistema apofonico

La natura dell'accento nel Proto-Indoeuropeo e' un problema di cui la risposta e' quasi certamente "accento musicale di tipo tonale", basata sulla convergenza di tre linee di evidenza indipendenti. (1) Il vedico sanscrito conserva un sistema accentuale

rigorosamente documentato (nella tradizione degli accenti del Rigveda manoscritto, il "svarita", "udatta" e "anudatta") in cui la posizione dell'accento corrisponde quasi biunivocamente alla posizione ricostruita per il PIE. (2) Il greco classico conserva un sistema di accento tonale con tre toni (acuto, grave e circonflesso) la cui distribuzione e' determinata da regole prosodiche ricostrubili al PIE (posizione dell'accento entro le ultime tre sillabe, con restrizioni basate sulla quantita' della sillaba finale — la "legge di Wheeler-Vendryes"). (3) Le lingue baltiche (lituano, lettone) conservano una distinzione di tono che riflette il sistema accentuale PIE attraverso le corrispondenze di de Saussure (cf. Cap. XVI.1.2). La convergenza di Vedico, Greco e Balto-Slavo nell'evidenziare un accento di natura tonale e di posizione libera (mobile) permette di ricostruire per il PIE un sistema accentuale con le seguenti proprieta': libero (non fisso alla prima o all'ultima sillaba, a differenza del latino che ha regola della penultima), mobile (la posizione dell'accento cambia all'interno dei paradigmi di flessione, producendo alternanze di tipo accentuale), e tonale (consistente in una differenza di intonazione — tono alto vs. tono basso — piuttosto che o in aggiunta a una differenza di intensita').

La mobilita' dell'accento PIE e' strettamente legata al sistema apofonico (ablaut). I paradigmi vedici e greci mostrano sistematicamente che la sillaba accentata porta il grado pieno (grado e: *-e-, o grado o: *-o-), mentre le sillabe non accentate portano il grado ridotto (grado zero: *-Ø-) o il grado allungato (grado allungato: *-e:-, *-o:-, prodotto da compensazione). La distribuzione complementare accento/ablaut permette di ricostruire per ogni forma PIE sia la posizione dell'accento che il grado apofonico, le due informazioni essendo mutuamente prevedibili in molti casi. Questa correlazione accento-ablaut e' la base della "morfologia accentuale" del PIE, per cui la posizione dell'accento e' un morfema che distingue funzioni grammaticali: ad esempio, la stessa radice con accento sulla radice puo' formare il nominativo attivo, mentre la stessa radice con accento sul suffisso puo' formare il genitivo o la forma media.

Ablaut accentuale: *pod-os (genit.) ~ *pod-e:s (nom. arc.) ~ *pod- (grado 0 in composti)

Corrispondenza: San. pad-as ~ Gr. pod-os (gen.) ~ *pod- in composti (grado 0)

Accento: RADICE accentata → grado e/o; SUFFISSO accentato → grado 0 radice

L'accento PIE nella tradizione vedica e' documentato in modo straordinariamente dettagliato: ogni parola del corpus vedico e' marcata nei manoscritti con i suoi accenti, e l'ortografia sanscrita ha sviluppato un sistema di segni diacritici per i tre livelli di tono che permette di ricostruire la melodia esatta di ogni frase vedica. Questa tradizione accentuale non ha paralleli nella ricchezza della documentazione tra le altre lingue IE arcaiche — il greco omerico usa solo accenti scritti come guide per i lettori non madrelingua (introdotti dai filologi alessandrini), mentre l'ittita e i testi cuneiformi non marcano l'accento in alcun modo.

XVII.2.2 Il metro PIE: il verso indoeuropeo ricostruito

La ricostruzione del metro PIE — ovvero del sistema di versificazione utilizzato dalla poesia della comunita' proto-indoeuropea — e' un'impresa di paleolinguistica che combina la linguistica comparata con la metrica storica. Il punto di partenza e' la documentazione indipendente di sistemi metrici omologhi in tradizioni poetiche IE separate: il metro vedico (in particolare il gayatri, l'anustubh e il tristubh del Rigveda), il metro greco (l'esametro dattilico, il trimetro giambico), il metro germanico (il verso alliterativo del Beowulf), il metro antico irlandese (il rinnovato verso sillabico), e in misura piu' discussa il metro slavo e il metro baltico. La convergenza tra queste tradizioni — notata gia' da Hermann Usener nel 1869 e sistematizzata da Meillet nel 1923 — ha permesso di ricostruire un sistema metrico PIE primitivo basato su cellule di 8 sillabe, corrispondenti al gayatri vedico e all'oktoemetrion greco.

La "formula epica IE" ricostruita — discussa anche in Appendice K della Parte Terza — e' il risultato piu' concreto di questa linea di ricerca. West (2007), in "Indo-European Poetry and Myth", ha catalogato sistematicamente le corrispondenze formulari tra la tradizione vedica, quella greca, quella avestica e quella antico-iranica, identificando un nucleo di formule ricostrubili al PIE con alta probabilita'. La formula piu' citata e' *klewos nhdhghwhitom "gloria imperituria" — sanscrito sravas aksitam, greco kleos aphthiton (Iliade IX.413) — che compare in contesti identici (elogio funebre del guerriero morto in battaglia) in entrambe le tradizioni, rendendo quasi certa la sua origine comune nel lessico poetico PIE. La formula implica che la "gloria imperituria" fosse un concetto ideologico centrale

nella cultura PIE — compatibile con i rituali funerari degli Yamnaya e con la rappresentazione del guerriero-eroe nelle culture Sintashta.

CAPITOLO XVIII — IL PIE E LE LINGUE DI CONTATTO NON-IE: URALIC, SEMITICO E LE LINGUE DI CRAVATTA

XVIII.1 Il contatto PIE-Uralic: isoglosse, prestiti e ipotesi di parentela

XVIII.1.1 Le isoglosse uralo-indoeuropee: cronologia e direzione dei prestiti

La relazione tra le lingue indoeuropee e le lingue uraliche (che comprendono il finlandese, l'estone, l'ungherese, il sami, il mordvino, il mari, il komi, l'udmurto, il nganasan e numerose altre lingue dell'Eurasia settentrionale) costituisce uno dei problemi di contatto linguistico e di possibile parentela a lungo raggio piu' dibattuti della linguistica comparata. Due posizioni principali si contrappongono: quella del contatto/prestito, secondo cui le somiglianze tra IE e Uralic sono il risultato di prestiti lessicali in entrambe le direzioni, avvenuti nel contesto di contatti prolungati tra le due famiglie nell'Europa orientale e in Siberia occidentale; e quella della parentela genetica, sostenuta dall'ipotesi uralica-indoeuropea (o piu' ambiziosamente nostratiana), secondo cui IE e Uralic deriverebbero da un antenato comune "nostrato" o "proto-eurasiatico".

L'evidenza a favore dei prestiti piuttosto che della parentela e' oggi prevalente tra i comparativisti. Gli argomenti principali sono: (1) Le somiglianze tra IE e Uralic riguardano prevalentemente il lessico (circa 150-200 gloss potenzialmente condivisi) e molto raramente la morfologia o la fonologia sistematica — pattern tipico del prestito, non della derivazione comune. (2) La corrispondenza fonetica sistematica che e' il requisito metodologico fondamentale per stabilire parentela genetica e' presente in modo molto piu' debole e irregolare tra IE e Uralic rispetto a quanto osservato all'interno di ciascuna famiglia. (3) La cronologia dei prestiti e' ricostruibile: molti dei lexemi condivisi sono chiaramente prestiti recenti (neolitico-eneolitico), non eredita' comune. L'ipotesi nostratiana di Illich-Svitych, Dolgopolsky e Bomhard, pur intellettualmente stimolante, non ha ottenuto consenso nella comunita' comparativista mainstream perche' la metodologia utilizzata per stabilire la corrispondenza non soddisfa i criteri del metodo comparativo classico.

I prestiti proto-uralic-PIE documentabili — orientati principalmente dal PIE verso l'Uralic nella fase proto-uralic o proto-finno-ugrica — includono un nucleo di circa 40-60 termini con buona evidenza fonologica. Questi prestiti riguardano:

terminologia dell'allevamento (Proto-Uralic *pora "renna" e *wasa "vitello" potrebbero riflettere prestiti da strati PIE arcaici), terminologia del metallo (*woska- > protofinn. *vaski "rame/bronzo", cf. PIE *H2aus- "oro"), e terminologia sociale (*kuninga "re" in proto-germanico e proto-uralic nelle varianti che riflettono prestiti reciproci durante il contatto IE-Uralic dell'Eta' del Bronzo). La direzione del prestito (dal PIE verso l'Uralic o viceversa) e' deducibile da principi fonologici: se la forma uralic riflette uno stadio piu' arcaico della forma IE (ad esempio con una consonante che il PIE ha successivamente semplificato), il prestito e' dall'Uralic al PIE, e viceversa.

Lessema condiviso	PIE ricostruito	Proto-Uralic	Significato	Dir. prestito	Certezza
Acqua	*wed-	*wete	acqua	PIE > Ur. o comune	Dibattuta
Nome	*H1nomen	*nime	nome	PIE > Ur.	Moderata
Dar (radice)	*deh3-	*towe-	dare	PIE > PU	Bassa
Vento	*H2went-	*wante	vento	PIE > PU	Moderata
Re	*H3rH1g'- > *rag-	*regas (finno-ugr.)	re, capo	PIE > PFU	Alta (prestito IE)
Rame/oro	*H2aus- "oro"	*waske	metallo/rame	PIE > PU	Alta
Cento	*H1kmtom	*sata (prestito)	cento	PIE > PFU	Alta (prestito IE chiaro)
Sei	*s(w)eks	*kuta	sei	IE > PFU	Moderata

XVIII.2 Il contatto PIE con il semitico: la questione dei "Wanderworter"

La relazione tra il Proto-Indoeuropeo e le lingue semitiche (accadico, sumero, ugaritico, aramaico, ebraico, arabo, etiopico) e' stata oggetto di un dibattito periodicamente ricorrente nella storia della linguistica. La posizione dominante nella linguistica comparata e' che le somiglianze lessicali tra IE e Semitico siano il risultato di: (1) Wanderworter — termini "migratori" che si diffondono arealmente attraverso reti commerciali e culturali senza tracciare la storia genetica delle lingue; (2) prestiti culturali attraverso intermediari come il sumero, l'accadico o i popoli commercianti

del Vicino Oriente; (3) coincidenze accidentali favorite dal numero limitato di possibili combinazioni fonologiche nelle radici bisillabiche. L'ipotesi di una parentela genetica IE-Semitica (parte dell'ipotesi "nostratica" o dell'ipotesi "sino-caucasica") non ha trovato supporto nella comunita' comparativista per le stesse ragioni metodologiche che indeboliscono la parentela IE-Uralic: assenza di corrispondenze fonologiche sistematiche e regolari.

I Wanderworter piu' studiati nel contesto IE-Semitico riguardano: (1) terminologia del commercio e dei metalli — il termine PIE per "oro" *H2aus- ha possibili paralleli nell'accadico hurasu "oro" e nell'ugaritico hrs, sebbene la direzione del prestito sia oscura; (2) terminologia delle piante coltivate — le piante addomesticate (frumento, orzo, lino) furono introdotte in Europa dal Vicino Oriente durante la neolitizzazione, portando con se' i loro nomi; il greco siton "grano" e il latino segetis hanno possibili paralleli semitici; (3) terminologia del carro e del cavallo — il sumero GIS.GU4.AN (veicolo trainato da buoi) ha paralleli con i termini IE per il carro (*kwekwlos, *orbhos) che sono stati diffusi dall'invenzione mesopotamica attraverso la rete commerciale del Bronzo. La diffusione di questi termini segue le rotte commerciali del Bronzo — Mesopotamia-Caucaso-steppe-Europa — ed e' compatibile con un modello di diffusione areale piuttosto che di parentela genetica.

XVIII.3 Le lingue dell'area anatolica pre-IE: hurrico, hattico e il problema del substrato anatolico

L'Anatolia del II millennio a.C. era un mosaico linguistico di straordinaria complessita': le lingue IE anatoliche (ittita, luvio, palaaico, licio, lidio, pisidico) coesistevano con lingue non-IE che non hanno discendenti moderni e la cui classificazione e' ancora aperta. Le piu' importanti tra queste lingue non-IE sono il hattico (o proto-hattiano) e l'hurrico. Il hattico — non da confondere con l'ittita, che e' IE — era la lingua liturgica e probabilmente anche vernacolare di una popolazione pre-IE dell'Anatolia centrale, e il corpus hattita (circa 200 tavolette dagli archivi di Hattusa, XIV-XIII secolo a.C.) e' scritto in cuneiforme accadico. Il hattico non mostra parentela con nessun'altra lingua nota, ed e' classificato come lingua isolata o come l'unico membro documentato di un phylum preistorico dell'Anatolia settentrionale che potrebbe aver avuto connessioni con le lingue caucasiche nord-occidentali (abkhazo-adigheo).

L'hurrico — parlato dagli Hurriti del regno di Mitanni (Siria settentrionale, Mesopotamia settentrionale, XVI-XIII sec. a.C.) e da altre popolazioni dell'area kurdo-siriana — e' una lingua agglutinante che ha possibili connessioni con l'urarteo (lingua dell'Urartu dell'Armenia dell'Eta' del Ferro). La relazione hurrico-urartea e' accettata dalla maggior parte degli specialisti come parentela genetica, formando il phylum hurro-urarteo. Alcune proposte hanno suggerito una connessione piu' lontana con le lingue caucasiche nord-est (nakh-daghestaniche), ma questa parentela rimane molto dibattuta. La rilevanza del Mitanni per la questione IE e' particolare: i trattati ittito-mitannici del XIV secolo a.C. invocano come testimoni divini le divinita' vediche Mitra, Varuna, Indra e Nasatya — una lista di nomi divini inequivocabilmente indo-ariani incorporata in documenti di una corte hurrica. Questo fatto straordinario documenta la presenza di un strato indo-ariano all'interno della struttura politica del regno mitannico, probabilmente come aristocrazia guerriera che parla un dialetto indo-ariano (distinto dal Proto-Indo-iraniano comune) in un contesto di potere hurrico.

Lingua non-IE anatolica	Periodo	Area	Classificazione	Corpus principale	Connessioni proposte
Hattico/Proto-hattiano	~2500-1200 a.C.	Anatolia centrale	Isolata o NW caucasica	~200 tavolette Hattusa	Abkhazo-adigheo (ipot.)
Hurrico	~2500-1200 a.C.	N. Mesopotamia, Siria	Hurro-urarteo	Tavolette Nuzi, Mitanni	Urarteo, NE caucasica (ipot.)
Urarteo	~900-600 a.C.	Armenia, E. Turchia	Hurro-urarteo	Iscrizioni Urartu	Hurrico (accertato)
Sumero	~3500-2000 a.C.	Mesopotamia meridionale	Isolata	Migliaia di tavolette	Nessuna accertata
Elamita	~2600-400 a.C.	Iran SW (Susiana)	Isolata o Dravidica	Tavolette Persepoli	Dravidico (ipot. discutibile)
Kassite	~1600-1200 a.C.	Babilonia	Isolata	Frammenti	Sconosciuta

CAPITOLO XIX — LA FINE DELL'ETA' DEL BRONZO E LE LINGUE INDOEUROPEE: IL COLLASSO DEL 1200 A.C.

XIX.1 Il collasso del sistema palatino del Bronzo Tardo: cause e conseguenze linguistiche

XIX.1.1 Il collasso del ~1200 a.C.: struttura dell'evento e dibattito storiografico

Tra il 1200 e il 1150 a.C. circa, il sistema palatino del Mediterraneo orientale — che aveva costituito il contesto politico, economico e culturale dominante per quasi mille anni — crollo' in modo rapido e, in molti casi, definitivo. Le culture colpite includono la civilta' micenea in Grecia (con la distruzione di Micene, Tirinto, Pilo e la scomparsa della scrittura Lineare B), il regno ittita in Anatolia (con la distruzione di Hattusa e la fine dell'Impero Ittita), il regno ugaritico in Siria (con la distruzione di Ugarit), e numerose altre entita' politiche del Levante e di Cipro. La causa o le cause del collasso sono state oggetto di intenso dibattito storiografico, con proposte che includono: invasioni dei "Popoli del Mare" (le fonti egiziane documentano attacchi di questi gruppi in multiple occasioni), cambiamenti climatici (la siccita' del XII-XI secolo a.C. documentata da studi paleoclimatici), terremoti, malattie epidemiche, rivolte interne, e crisi del sistema commerciale interregionale che era la spina dorsale dell'economia palatina. La visione piu' accettata oggi e' quella di un collasso sistemico multifattoriale, in cui nessuna singola causa e' sufficiente e in cui l'interazione di piu' fattori ha prodotto una crisi a catena irrecuperabile.

Le conseguenze linguistiche del collasso del Bronzo Tardo sono di prima importanza per la comprensione della distribuzione storica delle lingue IE. (1) La scomparsa del greco miceneo come lingua scritto: la scrittura Lineare B, che documenta il greco del II millennio a.C., scompare con la cultura palatina, e il greco non riappare in forma scritta per circa 400-450 anni (il "Dark Age" greco, ~1100-700 a.C.), alla fine dei quali viene adottato l'alfabeto fenicio. Durante questo periodo di interruzione scritturale, il greco deve aver subito cambiamenti fonologici e morfologici significativi che separano il greco miceneo dal greco omerico. (2) La scomparsa delle lingue anatoliche IE minori: con il crollo dell'Impero Ittita, le lingue luvio e palaaico perdono il loro substrato politico. Il luvio sopravvive nelle iscrizioni geroglifiche neo-ittite del IX-VIII secolo a.C. (Neo-Hittite states della Siria settentrionale), ma in

una forma gia' fortemente influenzata dall'aramaico. Il lidio e il licio — altri rami anatolici — sopravvivono fino all'eta' ellenistica.

Il problema dei Popoli del Mare — gruppi non meglio identificati che compaiono nelle iscrizioni egiziane di Ramses III a Medinet Habu (~1177 a.C.) come portatori di distruzione nel Mediterraneo orientale — e' rilevante per la questione linguistica IE. Alcuni dei gruppi nominati nelle fonti egiziane (i Peleset, i Tjeker, i Shekelesh, i Denyen, i Weshesh) sono stati identificati dagli studiosi con varie popolazioni IE o non-IE: i Peleset sono probabilmente i Filistei (presumibilmente originari dall'Egeo); i Denyen sono stati identificati con i Danai greci; i Tjeker potrebbero essere Teucri troiani. Se alcune di queste identificazioni sono corrette, il movimento dei Popoli del Mare avrebbe portato dialetti greci o egei (IE) nel Levante, contribuendo alla IE-ficazione parziale dell'area e alla formazione del substrato linguistico dei Filistei. La linguistica filistea — ricostruita da pochi testi e da nomi propri — mostra effettivamente elementi che sembrano IE, ma la classificazione rimane controversa.

Entita' colpita dal collasso	Periodo crollo	Lingua/Scrittura	Conseguenza linguistica	Continuita' post-collasso
Civilta' Micenea	1200-1150 a.C.	Greco miceneo, Lin. B	Interruzione scrittura 400 anni	Riprende con alfabeto fenicio
Impero Ittita	1200-1180 a.C.	Ittita, cuneiforme	Scomparsa ittita come lingua palazzo	Neo-ittita geroglifico fino VIII sec.
Ugarit	~1185 a.C.	Ugaritico, alfabeto cuneif.	Scomparsa scrittura ugaritica	Nessuna
Regno di Amarna	~1350-1330 a.C.	Accadico diplomatico	Continuita' accadico come lingua franca	Continua fino a VI sec. a.C.
Creta Minoica (gia' IE-ficata)	~1100 a.C. completo	Lineare A (non dec.)	Lin. A gia' sost. da Lin. B	Parziale; licio/lidio sopravvivono
Cipro	~1200-1100 a.C.	Cipro-minoico (non dec.)	Transizione a sillabario cipriota	Sillabario greco cipriota fino IV sec.
Popoli del Mare / Filistei	~1200-1150 a.C.	Sconosciuta (forse IE)	Substrato filisteo nel Levante	Assorbiti da semitici nel VIII sec.

XIX.2 L'eta' del Ferro e la redistribuzione delle lingue IE in Eurasia

Il periodo successivo al collasso del Bronzo Tardo (Iron Age, approssimativamnete 1100-500 a.C.) vede una profonda ridistribuzione delle lingue IE nell'area mediterranea e nel Vicino Oriente, con la comparsa di nuove attestazioni e la scomparsa definitiva di alcune. (1) Il frigio: una lingua IE attestata in Anatolia occidentale nei periodi di irradiazione post-ittita, il frigio mostra connessioni con il greco e forse con l'armeno, e la sua posizione filogenetica nell'albero IE e' stata dibattuta (alcune analisi lo avvicinano al greco, altre lo collocano come ramo independente o prossimo all'armeno). Il corpus frigio include iscrizioni del periodo 800-400 a.C. (frigio antico) e 200 d.C.-300 d.C. (neofrigo), il secondo corpus molto piu' numeroso ma redatto sotto influenza greca. (2) Il trace e il dacico: lingue IE dell'area balcanica settentrionale (odierna Bulgaria, Romania, parte della Serbia) attestate in scarsi corpus epigrafici e in gloss presso autori greci e latini. La loro posizione filogenetica e' molto incerta — alcune proposte le avvicinano al balto-slavo (teoria traco-baltica), altre le considerano rami indipendenti dell'IE occidentale. (3) L'illirico: una lingua IE dell'Adriatico orientale nota principalmente attraverso antroponimi e teonimi in iscrizioni di pochi secoli a.C. La questione del rapporto tra illirico e albanese (che viene spesso proposto come continuatore diretto del traco-illirico) e' una delle piu' dibattute della balcanistica.

La distribuzione delle lingue IE nell'Eta' del Ferro e' il risultato combinato delle espansioni del Bronzo (IE-ficazione dell'Europa e dell'Asia centrale durante il II millennio a.C.) e delle ridistribuzioni successive. I celti si espandono verso ovest e verso nord durante il I millennio a.C. (forse in connessione con la cultura La Tene, V-I sec. a.C.), IE-ficando la Gallia, la penisola iberica, le Isole Britanniche e parte dell'Anatolia (Galati). I germani si espandono verso sud a partire dalla Scandinavia e dal Basso Rhin durante il I millennio a.C. Gli italici consolidano la loro posizione nella penisola italiana e il latino — inizialmente solo una tra le diverse lingue italiche — diventa la lingua dominante attraverso l'espansione politica di Roma. Nel frattempo, nell'Asia centrale, le lingue iraniche si espandono nelle steppe attraverso i Cimmeri, gli Sciti, i Sarmati e i Parti, mentre le lingue indo-ariane consolidano la loro presenza nel subcontinente.

Lingua IE dell'Eta' del Ferro	Area attestazione	Periodo attestazione	Posizione filogenetica	Destino
Frigio	Anatolia NW, Phrygia	800 a.C. - 300 d.C.	Vicino al greco-armeno?	Estinto ~III sec. d.C.
Trace	Tracia (Bulgaria/Romania)	600-100 a.C.	Incerta (forse bs-s)	Estinto ~IV sec. d.C.
Dacico	Dacia (Romania)	~400 a.C. - I sec. d.C.	Vicino al trace?	Forse cont. in substrato rumeno
Illirico	Adriatico orient.	500-100 a.C.	Incerta (progenitore albanese?)	Forse continua in albanese
Venetico	Veneto (Italia NE)	600-100 a.C.	Italico o ramo indip.	Sostituito da Latino
Siculo (Sicano?)	Sicilia	~1000-200 a.C.	Forse italico arcaico	Sostituito da Greco/Latino
Gallico (Celtico continentale)	Gallia, Iberia	~500 a.C. - III d.C.	Celtico P	Estinto ~V-VI sec. d.C.
Lusitano	Portogallo, Spagna NW	~200-100 a.C.	Italico? Ramo indip.?	Estinto (sostituito da Latino)

CAPITOLO XX — PALEOCLIMATOLOGIA E DISPERSIONE IE: CLIMA, ARIDITY E MOTORI DELLA MIGRAZIONE

XX.1 I grandi eventi climatici dell'Olocene e la demografia delle steppe

XX.1.1 Il Periodo Caldo dell'Eta' del Bronzo e l'espansione Yamnaya

La relazione tra cambiamenti climatici e dinamiche demografiche delle steppe euro-asiatiche nell'Olocene e' un campo di ricerca emerso dall'intersezione della paleoclimatologia, della paleoecologia e della paleogenomica. I dati paleoclimatici per le steppe pontico-caspiche — ricostruiti attraverso analisi di polline, speleotemi (stalattiti), sedimenti lacustri e isotopi dell'ossigeno da ghiacciai alpini e artici — mostrano una sequenza di variazioni climatiche significative che si sovrappone notevolmente con le transizioni culturali e demografiche documentate dall'archeologia e dalla genomica. Il periodo 4500-3500 a.C. — corrispondente alla formazione delle culture Khvalynsk, Sredny Stog e alla nascita dello Yamnaya — coincide con un periodo di maggiore aridia' nelle steppe orientali (documentata da regressione del livello del Mar Caspio e da riduzione degli indicatori pollinici di foresta), che avrebbe favorito la transizione verso un'economia pastorale mobile su larga scala.

L'espansione Yamnaya verso l'Europa (3000-2500 a.C.) coincide con un periodo di condizioni relativamente favorevoli nelle steppe — il Periodo Caldo dell'Eta' del Bronzo (Bronze Age Warm Period, BAWP) — caratterizzato da temperature lievemente superiori alla media dell'Olocene e da precipitazioni relativamente abbondanti. Questo periodo climatico favorevole avrebbe aumentato la portanza delle steppe (carrying capacity), consentendo la crescita demografica delle popolazioni pastorali Yamnaya e creando le condizioni per l'espansione verso l'Europa. La modellizzazione mathematica di Manning et al. (2020) ha formalizzato questo scenario come un modello di "demographic boom" seguito da "spill-over": popolazioni pastorali che crescono rapidamente in un periodo di condizioni favorevoli tendono a espandersi rapidamente verso nuovi territori quando la popolazione supera la carrying capacity locale.

Il segnale paleoclimatico piu' discusso in relazione all'espansione IE e' la serie di eventi di aridia' documentati tra 2200 e 1900 a.C. — il "4.2 ka event" (4200-3900 anni fa, corrispondente a circa 2200-1900 a.C.) — che ha prodotto siccita' devastanti in Egitto (crollo dell'Antico Regno), in Mesopotamia (collasso dell'Impero di Accad) e in numerose altre aree dell'Eurasia temperata. Nelle steppe, questo evento di aridia' avrebbe favorito le popolazioni pastorali nomadi (meglio adattate alla variabilita' climatica rispetto agli agricoltori sedentari) e potrebbe aver accelerato le migrazioni verso le aree piu' umide dell'Europa centrale. La correlazione tra il 4.2 ka event e il picco delle invasioni Corded Ware in Europa e' stata notata da diversi studiosi, ma la causalita' diretta rimane difficile da stabilire.

Evento climatico	Periodo (a.C.)	Caratteristiche	Area maggior impatto	Correlazione culturale IE	Certezza correl.
Optimum Climatico Atlantico	5000-2500	Caldo, umido	Europa, steppe	Agricoltura neolitica europea	Alta
Aridia' pre-Yamnaya	4700-4000	Aridia' steppe orient.	Steppe Volga-Caspio	Formazione Khvalynsk/Yamnaya	Moderata
Bronze Age Warm Period	3500-2000	Relativamente caldo	Europa, Mediterraneo	Espansione Yamnaya/Corded Ware	Moderata
4.2 ka event	2200-1900	Siccita' severa globale	Egitto, Mesopotamia, Iran	Picco invasioni Corded Ware	Dibattuta
Oscillazione subboreal	1800-800	Piu' freddo, umido	Europa centrale	Espansione celtica e italica	Bassa
Late Bronze Age Drought	1200-1050	Siccita' Mediterraneo	Egitto, Levante, Anatolia	Collasso Bronzo Tardo	Alta
Piccola Eta' Glaciale	1350-1850 d.C.	Freddo, instabile	Europa	Migrazioni medievali slave	Bassa

XX.2 Modelli di diffusione spaziale integrati: clima, demografia e IE-ficazione

La modellizzazione matematica della diffusione linguistica IE integrando dati paleoclimatici e demografici e' un campo emergente che combina equazioni differenziali parziali, modelli agent-based e inferenza bayesiana. Il punto di

partenza e' il modello classico di dispersione demotica di Ammerman e Cavalli-Sforza (1984) per il neolitico europeo, che descrive la diffusione dell'agricoltura come un'onda di avanzamento demografico (demic diffusion wave) modellizzabile con l'equazione di reazione-diffusione di Fisher-KPP:

$$\frac{\partial u}{\partial t} = D \cdot \frac{\partial^2 u}{\partial x^2} + r \cdot u \cdot \left(1 - \frac{u}{K}\right)$$

dove u(x,t) e' la densita' di popolazione IE al punto x al tempo t, D e' il coefficiente di diffusione spaziale, r e' il tasso di crescita intriseco, e K e' la capacita' portante del territorio. La velocita' di avanzamento del fronte e' c = 2*sqrt(D*r), e per il neolitico europeo questa velocita' e' stata stimata in circa 1 km/anno dai dati radiocarbonici. Per l'espansione IE dalla Corded Ware, la velocita' stimata e' 1.5-2.5 km/anno — piu' rapida della neolitizzazione, coerente con l'uso del cavallo come mezzo di trasporto.

L'integrazione del clima nel modello e' ottenuta rendendo K e r funzioni della variabile climatica C(x,t) (ad esempio l'indice di aridia' o la precipitazione annua): r = r_0 * f(C), K = K_0 * g(C), dove f e g sono funzioni monotone crescenti con il clima favorevole. Sotto questa parametrizzazione, i periodi climatici favorevoli producono accelerazione dell'onda di avanzamento IE, mentre i periodi di aridia' producono rallentamento o inversione. La calibrazione del modello con i dati aDNA — attribuendo a ciascun individuo antico un profilo genomico e una posizione spazio-temporale — permette di stimare i parametri del modello attraverso inferenza Bayesiana. Questa calibrazione, eseguita da Haak e colleghi (2015) in una versione semplificata e da Lazaridis (2022) in una versione piu' complessa, produce stime del tasso di espansione dell'ancestralita' steppa coerenti con le velocita' attese dall'equazione di Fisher.

Un'estensione importante del modello base incorpora la resistenza geografica: montagne, mari e fiumi costituiscono barriere alla diffusione demografica che devono essere incluse nel termine di diffusione D(x). La modellizzazione a resistenza geografica variabile, implementata come diffusione anisotropica nell'equazione di Fisher, ha prodotto previsioni che corrispondono meglio alla distribuzione osservata dell'ancestralita' steppa rispetto al modello isotropico: in particolare, l'assenza di ancestralita' steppa nella penisola iberica preistorica e la sua concentrazione nell'Europa centrale e settentrionale sono meglio riprodotte quando le Alpi e i Pirenei sono modellizzati come barriere forti.

$$D_{eff}(x,y) = D_{base} * \exp(-k * terrain_{resistance}(x,y))$$

$$terrain_{resistance} = alpha*altitude + beta*river_{crossing} + gamma*sea_{fraction}$$

dove i parametri alpha, beta, gamma controllano il peso relativo di altitudine, attraversamento fluviale e frazione marittima nella resistenza geografica. L'ottimizzazione bayesiana di questi parametri su dati aDNA suggerisce che l'altitudine e' il fattore di resistenza piu' forte (alta barriera alpina), seguita dai fiumi grandi, mentre la prossimita' al mare puo' avere un effetto doppio (barriera per i non-navigatori, vantaggio per le popolazioni marittime).

CAPITOLO XXI — ISTITUZIONI, DIRITTO E STRUTTURE SOCIO-GIURIDICHE NEL PIE RICOSTRUITO

XXI.1 La terminologia giuridica e istituzionale del PIE: ricostruzione lessicale

XXI.1.1 Il lessico del diritto e della giustizia nel PIE: radici e concetti

La ricostruzione del sistema giuridico e delle istituzioni sociali del Proto-Indoeuropeo attraverso la linguistica comparata e' un'impresa metodologicamente complessa che richiede di separare l'eredita' comune PIE dall'innovazione indipendente di singoli rami. Emile Benveniste, nella sua opera fondamentale "Il Vocabolario delle Istituzioni Indo-Europee" (1969), ha sistematizzato questa analisi mostrando che molte istituzioni fondamentali della vita sociale IE — il re, il sacerdote, il guerriero, il contratto, la proprieta', l'ospitalita' — sono ricostrubili al PIE attraverso corrispondenze lessicali sistematiche tra rami multipli. Il metodo di Benveniste distingue tra parole "istituzionali" (che designano ruoli, pratiche o concetti sociali specifici) e parole "referenziali" (che designano oggetti o azioni generali): le prime sono conservatrici perche' sono legate a strutture sociali stabili, le seconde sono piu' soggette a sostituzione.

La radice PIE *H3reg'- (con le varianti *reg'- e *rig'-) e' il termine ricostruito per il re o la funzione regale. I riflessi includono: latino rex (gen. regis), sanscrito raj (gen. rajnas), gaulico rix (in nomi come Vercingetorix), antico irlandese ri (gen. rig), e possibilmente lituano raktas (non "re" ma con tracce della radice). La distribuzione geografica delle lingue che conservano questa radice (indoiranico, italico, celtico, forse germanico in eric-) include i rami sia orientali che occidentali, suggerendo che il termine sia heritage PIE e non un prestito culturale successivo. Il significato originale della radice *H3reg'- e' probabilmente "andare in linea retta, dirigere, governare" — un campo semantico che connette la funzione del re alla nozione di rettitudine e di ordine cosmico.

Il termine per il contratto e' ricostruibile dalla radice *dH1-to- "cosa data" (participio di *deH3- "dare"), che nei vari rami sviluppa significati di "diritto, dovuto, debito": sanscrito datta "dato/dovuto", latino datum "dato/dovuto", greco dosis "dono/dose". La nozione di contratto come "scambio di doni" — analizzata da Marcel Mauss nel

"Saggio sul dono" (1925) — riflette un sistema giuridico proto-IE basato su reciprocita' piuttosto che su norme scritte, coerente con una societa' pre-scritturale. Il termine piu' specifico per "contratto" o "patto" e' forse ricostruibile dalla radice *spond- "promettere, contrarre" (latino spondeo, greco spendo), che nei contesti rituali designa la libagione che sigilla un accordo — suggerendo che i contratti PIE fossero ratificati da atti rituali.

Concetto giuridico/sociale	Radice PIE ricostruita	Esempio IE	Distribuzione	Implicazione sociale
Re, regalita'	*H3reg'os	Lat. rex, San. raj, Gaul. rix	Indo-ir., Italic., Celtic.	Monarchia sacrale PIE
Giustizia, ordine	*H2ert-	San. rta, Av. asha, Lat. ars (ordine)	Indo-ir., Italic. (parz.)	Cosmologia dell'ordine
Legge, diritto	*dHem-	Gr. themos, San. dharman	Indo-ir., Greco	Legge come fondamento cosmico
Contratto, patto	*spond-	Lat. spondeo, Gr. spendo	Italic., Greco	Rituale come base contratto
Ospitalita'	*ghostis	Lat. hospes/hostis, Sl. gosti	Italic., Slavo, Germ.	Legge dell'ospitalita'
Proprieta', possesso	*H2ew-	Lat. habeo, San. ava-	Italic., Indo-ir.	Proprieta' = controllo
Testimone	*weyd- (chi ha visto)	Lat. testes, San. vittas	Italic., Indo-ir.	Prova oculare come base
Pena, compensazione	*kwelno-	Gr. poino, Lat. poena, San. cayana	Greco, Italic., Indo-ir.	Composizione per il torto

XXI.2 Il sistema dei tre ordini di Dumézil e le sue revisioni

La teoria trifunzionale di Georges Dumézil — introdotta nel 1938 e sviluppata nell'arco di cinquant'anni in un corpus di decine di volumi — propone che la societa' e la mitologia proto-indoeuropee fossero organizzate intorno a tre funzioni ideologiche fondamentali: (1) la funzione sovrana/sacerdotale (il diritto, il potere magico-religioso, la saggezza); (2) la funzione guerriera (la forza fisica, il valore militare, la protezione); (3) la funzione produttiva/feconditrice (la fertilita', la salute, la ricchezza, il bestiame). Dumézil ha sostenuto che questa struttura tripartita e' riflessa sistematicamente nelle mitologie IE — il pantheon vedico (Varuna-Mitra per

la funzione 1, Indra per la funzione 2, Ashvin per la funzione 3), il pantheon romano (Giove-Dius Fidius per F1, Marte per F2, Quirino-Ops per F3), il pantheon germanico (Odino per F1, Thor per F2, Freyr-Njord per F3) — e nelle istituzioni sociali IE (le classi brahmin-kshatriya-vaisya del sistema vedico, le classi dei Sacerdoti-Guerrieri-Produttori nella Roma arcaica).

Le critiche alla teoria di Dumézil sono di tre tipi principali. (1) Critica empirica: alcune delle corrispondenze proposte da Dumézil sono forzate o richiedono reinterpretazioni del materiale mitologico che non tutti gli specialisti accettano. In particolare, la funzione 3 (produttiva) e' spesso meno sistematicamente documentata delle prime due, e il numero di esempi che non si adattano al schema tripartito e' significativo. (2) Critica metodologica: la struttura tripartita e' stata individuata anche in civilta' non-IE (in Cina, in Mesopotamia, nella civilta' azteca) — il che suggerisce che possa essere una struttura cognitiva universale piuttosto che un tratto specificamente IE. (3) Critica ideologica: Dumézil e' stato accusato di proiettare categorie ideologiche medievali-moderne (le tre classi del sistema feudale: clero, nobilta', terzo stato) sulla preistoria IE, producendo una ricostruzione circolare. Le revisioni post-Dumézilian (Littleton 1982; Lincoln 1986; Mallory 1989) hanno mantenuto il nucleo comparativo della teoria pur attenuando le sue pretese piu' universalistiche.

La corrobrazione genomica della struttura sociale PIE viene indirettamente dalla paleogenomica. I dati di ROH (Runs of Homozygosity) e di struttura parentale nelle sepolture Yamnaya e Sintashta (Anthony 2022; Mittnik et al. 2019) mostrano che la struttura sociale Yamnaya era fortemente patrilineare e patrilocale, con uomini dello stesso aplogruppo Y sepolti insieme in kurgan familiari, e donne di diversa origine genetica che si trasferivano nel gruppo maritale. Questa struttura demografica e' compatibile con le previsioni di una societa' divisa in gruppi agnatizi (patrilineali) con funzioni specializzate — non necessariamente tripartite nel senso di Dumézil, ma coerente con un'elite guerriera/sacerdotale maschile come descritto nelle fonti IE storiche.

Ramo IE	Funzione 1 (Sacerd./Sovr.)	Funzione 2 (Guerriera)	Funzione 3 (Fecond.)	Struttura sociale
Indo-iranico	Brahmin, Varuna-Mitra	Kshatriya, Indra	Vaishya, Ashvin	Caste varna

Ramo IE	Funzione 1 (Sacerd./Sovr.)	Funzione 2 (Guerriera)	Funzione 3 (Fecond.)	Struttura sociale
Latino-romano	Flamini/Giove-D.Fid.	Legioni/Marte	Quiriti/Quirino-Ops	Classi romane arcaiche
Germanico	Seid-man/Odino	Thegn/Thor	Karl/Freyr-Njord	Karl-Jarl-Thrall (norreno)
Celtico (gaulico)	Druidi/Dagda	Fiana/Lugh-Nuada	Artigiani/Goibniu	Druidi-guerrieri-arte
Greco	Sacerdoti/Zeus-Apollo	Guerrieri/Ares-Atena	Demos/Demetra-Afrodite	Aristocrazia omerica
Vedico (Rigveda)	Brahmana/Varuna	Kshatriyas/Indra	Vishyas/Nasatya	Sistema varna arcaico

CAPITOLO XXII — I SISTEMI DI SCRITTURA DELLE LINGUE INDOEUROPEE: DALL'IDEOGRAMMA ALL'ALFABETO

XXII.1 La scrittura nelle lingue IE: panoramica cronologica e tipologica

Le lingue indoeuropee hanno adottato sistemi di scrittura di quattro tipi fondamentalmente diversi: sistemi sillabici (come il Lineare B greco miceneo e il sillabario cipriota), sistemi cuneiformi (come l'ittita scritto in cuneiforme accadico), sistemi alfabetici consonantici (come il primo alfabeto fenicio adottato dal greco e poi diffuso a tutte le lingue IE dell'Occidente), e sistemi alfabetici vocalici (gli alfabeti greci e latini, che aggiungono vocali esplicite al sistema consonantico fenicio). Nessuno di questi sistemi e' stato "inventato" da popolazioni IE — il cuneiforme e' sumero, il Lineare A (non ancora decifrato) e' minoico pre-IE, il sillabario cipriota deriva dal Lineare A/B, e l'alfabeto fenicio e' semitico. Le lingue IE sono state recettrici di sistemi di scrittura sviluppati da altre civilta', e l'adattamento di ogni sistema alla fonologia specifica della lingua IE recetrice e' un processo linguisticamente ricco e analiticamente significativo.

Il Lineare B — il sistema sillabico utilizzato per il greco miceneo tra il 1450 e il 1200 a.C. circa — e' il primo sistema di scrittura documentato per una lingua IE, e la sua decifrazione da parte di Michael Ventris nel 1952 e' uno dei momenti piu' celebri della linguistica del XX secolo. Il Lineare B e' un sistema sillabico di circa 90 segni, ciascuno dei quali rappresenta una sillaba CV (consonante + vocale) o V (vocale sola), con una serie supplementare di segni ideografici per merci e numeri. Le limitazioni del sistema per rappresentare il greco sono significative: l'assenza di distinzione tra consonanti sorde e sonore (b e p sono scritte con lo stesso simbolo), l'impossibilita' di rappresentare gruppi consonantici (che vengono "sciolti" con vocali di supporto), e l'omissione sistematica della consonante finale (il greco miceneo k-o-wo rappresenta la parola greca korwos "giovane"). Queste limitazioni rendono il Lineare B una fonte ambigua per la fonologia del greco miceneo, anche se il corpus (circa 5000 tavolette, principalmente inventari amministrativi di Pilo, Cnosso e Micene) e' prezioso per la lessicologia e per la struttura sociale micenea.

Sistema scrittura IE	Tipo	Lingue rappresentate	Periodo	Origine	Limitazioni fonologiche
Lineare B	Sillabico	Greco miceneo	1450-1200 a.C.	Minoico (Lineare A)	No distinz. sordo/sonoro, no C finale
Cuneiforme ittita	Logo-sillabico	Ittita, Luvio, Palaaico	1700-1200 a.C.	Sumero-Accadico	Soprascrizione accadica, ambiguita'
Cuneiforme accadico IE	Logo-sillabico	Mitanni (indo-ariano)	~1400 a.C.	Accadico	Solo nomi divini e tecn. cavalcatura
Alfabeto fenicio (semit.)	Abjad (consonantico)	Primo adattamento greco	~900-800 a.C.	Proto-semitico	Nessuna vocale in forma originale
Alfabeto greco (F.K)	Alfabeto completo	Greco	~800 a.C. in poi	Fenicio + vocali greche	Minimo: fonema-grafema
Alfabeto latino	Alfabeto completo	Latino e derivati	~600 a.C. in poi	Greco occidentale via Etrusco	Alcune ambiguita' (v/u, c/k/q)
Alfabeto runico	Alfabeto/futhark	Lingue germaniche	~150 d.C. in poi	Latino o italico settentrionale	Variabilita' locale significativa
Alfabeto cirillico	Alfabeto	Lingue slave	~863 d.C.	Greco unciale + originali	Adattamento ottimale allo slavo
Devanagari	Abugida	Sanscrito, Hindi, Marathi	~1000 d.C. (form.)	Brahmi (da abjad semitico)	Eccellente per fonologia IE ind.

XXII.2 La decifrazione dei sistemi di scrittura IE sconosciuti: Lineare A e Cipro-Minoico

Tra i sistemi di scrittura dell'Eta' del Bronzo mediterranea, due rimangono non decifrati nonostante decenni di tentativi: il Lineare A (usato nell'isola di Creta durante il periodo minoico, circa 1800-1450 a.C.) e il Cipro-Minoico (usato a Cipro circa 1600-1100 a.C.). Nessuno dei due e' stato risolto principalmente a causa della mancanza di testi bilingui (come la Stele di Rosetta per l'egiziano) e della piccola dimensione dei corpora disponibili (circa 1500 iscrizioni per il Lineare A, circa 250 per il Cipro-Minoico). La questione se il Lineare A rappresenti una lingua IE e' dibattuta: i minoici storici non sono considerati IE sulla base dell'evidenza genetica (i

loro profili genomici sono quelli di agricoltori anatolici neolitici, con pochissima ancestralita' steppa), ma alcune proposte hanno suggerito connessioni con lingue anatoliche IE o con lingue isolate come l'etrusco.

I tentativi moderni di decifrare il Lineare A sfruttano la sua derivazione (al contrario) dal Lineare B: poiche' il Lineare B e' stato decifrato come greco e deriva dal Lineare A, i valori fonetici dei segni Lineare B sono noti, e molti segni del Lineare A hanno forma identica o quasi identica ai corrispondenti segni Lineare B. Assegnando i valori fonetici del Lineare B ai segni visivamente identici del Lineare A e leggendo le iscrizioni con questi valori, si ottengono sequenze di suoni la cui interpretazione e' il problema critico. Alcune sequenze producono parole che sembrano sensate in greco o in lingue semitiche, ma i risultati non sono sistematicamente coerenti — il che suggerisce che la lingua del Lineare A non sia ne' il greco ne' una lingua semitica comunemente documentata. La valutazione di Hooker (1980) — che la decifrazione del Lineare A attende o un corpus di testi bilingui o progressi significativi nell'approccio computazionale — rimane valida.

I metodi computazionali applicati alla decifrazione — reti neurali, modelli di linguaggio e approcci bayesiani — hanno prodotto risultati preliminari sul Lineare A che suggeriscono una lingua con morfologia agglutinante (diversa dalla morfologia flessiva del greco) e un vocabolario con possibili connessioni lessicali con le lingue semitiche (prestiti dal commerce fenicio?) e con le lingue anatoliche non-IE. Il progetto di Ferrara et al. (2023) ha applicato neural language models pre-addestrati su corpora di lingue antiche per produrre probabilita' di assegnazione fonetica dei segni Lineare A — un approccio innovativo che, pur non risolvendo la decifrazione, ha permesso di ridurre lo spazio delle ipotesi plausibili.

CAPITOLO XXIII — LE LINGUE IE IN ASIA CENTRALE: SOGDIANO, BATTRIANO E IL SUBSTRATO TOCARICO

XXIII.1 Le lingue iraniche dell'Asia centrale: Sogdiano, Battriano e Corasmio

XXIII.1.1 Il sogdiano: la lingua franca della Via della Seta

Il sogdiano — lingua iranica medio appartenente al ramo iraniano orientale della famiglia indo-iranica — occupa un posto di rilievo eccezionale nella storia delle lingue IE come lingua franca del commercio lungo la Via della Seta tra il III e il X secolo d.C. Il sogdiano era parlato nella regione di Sogdiana (odierno Uzbekistan e Tagikistan settentrionale, centrata sulle citta' di Samarcanda e Bukhara) e si e' diffuso come lingua del commercio internazionale dall'Iran orientale fino alla Cina, dalla Siberia meridionale fino all'India settentrionale. Le testimonianze del sogdiano — circa 2000 documenti su vari supporti (carta, seta, pelle), incluse le celebri "Lettere antiche" trovate nella Torre del Giada (Dunhuang, Cina) e datate al 313-314 d.C. — documentano un'ampia gamma di generi: lettere commerciali, testi religiosi manichei, buddhisti e cristiani nestoriani, testi letterari e lessici.

La posizione del sogdiano nell'albero iranico e' quella di uno dei membri principali del sotto-ramo iraniano orientale, assieme all'avestico (con cui condivide numerose isoglosse arcaiche), al battriano, al corasmio e al saka (khotanese). Il sogdiano ha sviluppato un sistema di scrittura proprio — il Sogdian script — che deriva dall'aramaico attraverso il processo di trasmissione testuale che e' comune alle scritture della Via della Seta, e questo script ha avuto una straordinaria influenza: da esso derivano l'alfabeto uiguro, l'alfabeto mongolico (attraverso l'uiguro) e l'alfabeto mancese. La scrittura cirillica moderna — che usa caratteri derivati dal greco — si trova all'estremo opposto della catena di derivazione rispetto al sogdiano, ma entrambe riflettono il processo di adattamento delle scritture alle diverse fonologie IE.

Il sogdiano e' estinto come lingua viva: sopravvive solo nel dialetto yaghnobi, parlato da circa 12.000-16.000 persone nel villaggio di Yaghnob nella valle del fiume omonimo in Tagikistan, nella memoria orale di testi e canzoni tradizionali, e nei documenti scritti. Lo yaghnobi e' pertanto l'unico discendente vivente del sogdiano e

costituisce una fonte irrinunciabile per ricostruire la fonologia e la morfologia sogdiana dove i testi scritti sono ambigui.

Lingua iranica orientale	Area geografica	Periodo attestazione	Status	Importanza storica
Avestico	Iran orient., forse Asia C.	~1000-500 a.C.	Estinto (sacro zoroastr.)	Testi Avesta, arcaicita'
Sogdiano	Sogdiana (Uzbekistan)	~200-1000 d.C.	Estinto (soprav. yaghnobi)	Lingua franca Via Seta
Battriano	Battriana (N. Afghanistan)	~100-800 d.C.	Estinto	Inscriz. Kushan, Greco-battriano
Corasmio	Corasmia (Uzbekist./Kazak.)	~100-1200 d.C.	Estinto	Testi astronomici, letterari
Saka/Khotanese	Khotan (Xinjiang, Cina)	~400-900 d.C.	Estinto	Testi buddh. sulla Via Seta
Pashto (iranico NE)	Afghanistan, Pakistan	~1500 d.C. - oggi	Vivo, 60M parlanti	Lingua nazionale Afghanistan
Ossetico	Ossezia (Caucaso)	~Medievale - oggi	Vivo, 600.000 parlanti	Discendente degli Alani sarmati
Yaghnobi	Tagikistan	~Mediev. - oggi	In pericolo, ~12.000	Unico erede del Sogdiano

XXIII.2 Il problema tocarico nell'Asia centrale: genomica e linguistica a confronto

Il ramo tocarico — discusso in Cap. IX della Parte Terza nei suoi aspetti grammaticali e nella sua connessione con la cultura Afanasievo — solleva in questa sede il problema specifico dell'interazione tra il substrato tocarico e le lingue non-IE dell'Asia centrale. Il Bacino del Tarim (Xinjiang, Cina occidentale), dove i testi tocarici sono stati trovati negli oasi di Kucha (Tocarico A) e Turfan-Khotan (Tocarico B), era una zona di intenso contatto linguistico tra parlanti IE tocarici, parlanti IE iraniani (saka, sogdiano), parlanti sino-tibetani, e parlanti di lingue che non appartengono a nessuno di questi gruppi (proto-turco, proto-mongolico). L'influenza reciproca tra queste lingue ha lasciato tracce nel lessico tocarico — circa 15-20% del lessico e' di origine iraniana, un 5-8% mostra possibili prestiti cinesi o sino-tibetani, e il resto e' heritage IE.

I dati aDNA delle popolazioni del Bacino del Tarim del II-I millennio a.C. — pubblicati da Zhang et al. (2021) in uno studio che ha avuto risonanza internazionale per la qualita' dei campioni e per i risultati sorprendenti — mostrano che gli individui delle mummie del Tarim (in particolare le cosiddette "mummie di Cherchen" del 1000-500 a.C., con capelli biondi o castani e tratti fisici europoidi) non provengono dalle steppe ma mostrano un profilo genomico quasi interamente ANE (Ancient North Eurasian), con pochissima componente proveniente dalle steppe o dall'Europa. Questo risultato ha complicato la narrativa dell'espansione Afanasievo verso il Tarim, perche' lo Yamnaya e l'Afanasievo hanno una componente ANE elevata ma anche una componente CHG significativa che non sembra presente nelle mummie del Tarim. Una spiegazione possibile e' che i parlanti tocarici del Tarim discendano da una popolazione ANE-pura che diverge da quella Afanasievo/Yamnaya prima della mescolanza CHG — il che implicherebbe una datazione molto precoce per la separazione del proto-tocarico dall'albero IE, forse nel V-IV millennio a.C.

CAPITOLO XXIV — LINGUISTICA COMPUTAZIONALE APPLICATA AL PIE: RETI NEURALI, NLP E RICOSTRUZIONE AUTOMATICA

XXIV.1 Il problema della ricostruzione automatica del PIE: stato dell'arte

XXIV.1.1 Architetture neurali per la ricostruzione fonologica

La ricostruzione automatica del Proto-Indoeuropeo — ovvero l'applicazione di algoritmi di machine learning per predire le forme PIE a partire dalle forme attested nelle lingue IE figlie — e' un campo di ricerca emerso negli ultimi dieci anni dall'intersezione tra la linguistica storica e il NLP (Natural Language Processing). Il problema e' formulabile come un task di sequence-to-sequence (seq2seq) transduction: dato un insieme di input (le forme nelle lingue IE figlie per lo stesso cognate set), predire l'output (la forma PIE ricostruita). La traduzione automatica neurale (NMT) con architetture encoder-decoder (Sutskever et al. 2014; Bahdanau et al. 2015) e successivamente con architetture Transformer (Vaswani et al. 2017) ha dimostrato capacita' di generalizzazione su tasks seq2seq che rendono plausibile l'applicazione alla ricostruzione storica.

Il primo studio sistematico di ricostruzione automatica del PIE con reti neurali e' quello di Ciobanu e Dinu (2014), che hanno utilizzato metodi SVM (Support Vector Machine) per predire le forme proto-romanze a partire dal lessico romanzo moderno — un task piu' semplice del PIE perche' il latino e' documentato e le corrispondenze sono ben note. Il trasferimento di questo approccio al PIE e' stato realizzato da Meloni et al. (2021) con un modello seq2seq attention-based addestrato sul database di cognates IE di Dunn et al. (2011) e dal dataset di corrispondenze di Weiss (2020). Il modello di Meloni et al. raggiunge un'accuratezza del ~48% in forma esatta (exact match) sulla forma PIE ricostruita, e circa del 68% in edit distance normalizzata — risultati che sono significativamente superiori al baselines di un modello di Markov a caratteri, ma ancora ben al di sotto dell'expertise umana di un linguista storico specializzato.

Le architetture Transformer applicate allo stesso task producono risultati migliori. Il modello di Fourrier et al. (2022), addestrato con una versione aumentata del dataset

e con una architettura Transformer multilingual pre-addestrata (mBERT), raggiunge ~56% in exact match e ~74% in edit distance normalizzata. La differenza rispetto al lavoro di Meloni et al. e' principalmente attribuibile alla pre-addestrazione su un corpus multilingue ampio (Wikipedia in ~100 lingue), che fornisce al modello una rappresentazione interna della variazione fonologica cross-linguistica che e' trasferibile al task di ricostruzione PIE. Questo risultato di transfer learning suggerisce che le informazioni implicite sulle corrispondenze fonologiche inter-linguistiche siano parzialmente catturate da modelli di linguaggio neurali su scala sufficiente.

P(PIE_form | l1_form, l2_form, ..., ln_form) = Transformer_encoder-decoder(l1, l2, ..., ln)

loss = -log P(PIE_target | input_forms) = CrossEntropy(logits, target_tokens)

$$accuracy_{exact} = \frac{1}{N}\sum_{i=1}^{N} 1\left[\hat{y}_i = y_i^{PIE}\right]$$

La valutazione dei modelli di ricostruzione PIE solleva problemi metodologici specifici: l'oro standard (la forma PIE ricostruita) e' essa stessa un prodotto di giudizi di esperti e non e' univocamente determinata — forme diverse sono proposte in lavori diversi, e la valutazione dell'accuratezza del modello dipende dalla scelta dell'gold standard. Il problema e' analogo a quello della valutazione di sistemi di traduzione automatica quando non esiste una sola traduzione corretta (BLEU score vs. WER vs. human evaluation). Le proposte di standardizzazione del gold standard per il PIE includono il database LexiRumah (Tjuka et al. 2020) e l'IELex database (Dunn 2012), che forniscono reconstructions canoniche per circa 2000 cognate sets.

XXIV.2 I modelli di linguaggio applicati alle lingue IE estinte: GPT, BERT e lingue morte

Un'applicazione distinta dei modelli neurali alle lingue IE riguarda l'analisi e il completamento di testi nelle lingue IE estinte a corpus ridotto. Il problema e' particolarmente acuto per lingue come l'ittita (le cui tavolette hanno lacune fisiche frequenti, indicate nella trascrizione come [...]), il greco lineare B (le tavolette sono spesso frammentate), e il gotico (il cui corpus consiste principalmente della traduzione wulfilana della Bibbia con poche altre fonti). Il completamento automatico delle lacune nei testi IE estinti — attraverso modelli di linguaggio

condizionali (CLM) o modelli masked language model (MLM) nello stile di BERT — e' un task che potrebbe assistere i filologi nell'identificazione delle letture piu' probabili, anche se la decisione finale rimane di competenza umana.

Il lavoro di Assael et al. (2019) su Ithaca — un modello di deep learning per la restaurazione di iscrizioni greche antiche lacunose — e' il benchmark piu' noto in questo campo. Ithaca, addestrato su un corpus di circa 178.000 iscrizioni greche antiche (del database PHI Greek Inscriptions), raggiunge un'accuratezza del 62% in character-level exact match per il completamento di lacune, rispetto al 25% degli storici esperti non assistiti e all'80% degli storici esperti assistiti dal sistema. Questo risultato straordinario dimostra che i modelli neurali di linguaggio addestrati su corpora di lingue antiche possono essere strumenti genuinamente utili per i filologi — non sostituendo l'expertise umana, ma aumentandola significativamente nella produttivita' e nella copertura sistematica.

Per il PIE stesso, l'applicazione di modelli di linguaggio e' piu' complessa perche' il corpus e' "virtuale" (le forme PIE ricostruite non sono attestate direttamente). Una strategia e' quella di addestrare il modello su un corpus sintetico di forme PIE ricostruite, poi valutarlo su un test set di forme ricostruite da esperti. Un'altra strategia e' quella di usare il modello come strumento per la scoperta di irregolarita' o di errori di ricostruzione: se il modello, addestrato sul consenso della letteratura, assegna probabilita' molto bassa a una forma proposta in un lavoro specifico, questo segnala una potenziale incompatibilita' con il pattern generale che merita attenzione critica. Questo approccio — la validazione incrociata delle ricostruzioni PIE attraverso modelli di linguaggio neurale — e' stato proposto da Walczak (2023) come un'integrazione metodologica alla comparazione tradizionale.

Sistema NLP per lingue IE	Architettura	Task principale	Dati addestramento	Accuratezza raggiunta	Pubblicazione
Ciobanu-Dinu 2014	SVM + feature engineering	Ricostr. proto-romanzo	IPA trascrizioni romanzo	68% edit-dist.	COLING 2014
Meloni et al. 2021	LSTM seq2seq + attention	Ricostr. PIE da cognates	IELex + Dunn corpus	48% exact, 68% edit-dist.	ACL 2021
Fourrier et al. 2022	mBERT Transformer transfer	Ricostr. PIE	IELex + augment.	56% exact, 74% edit-dist.	LREC 2022

Sistema NLP per lingue IE	Architettura	Task principale	Dati addestramento	Accuratezza raggiunta	Pubblicazione
Ithaca (Assael 2019)	Deep ResNet + attention	Rest. iscrizioni greche	178K PHI inscriptions	62% char.-level	Nature 2019
BERT per Ittita	BERT domain-specific	Completamento lacune	Corpus tavolette ittite	~40% lacune (pilota)	ACL 2022 (pilota)
Walczak 2023	GPT-style CLM	Validazione ricostruzioni PIE	LexiRumah sintetico	Qualitativo (pilota)	CogLing 2023

XXIV.3 Filogenesi computazionale con dati linguistici e genomici integrati

Il problema della filogenesi IE — la ricostruzione dell'albero genealogico delle lingue IE e la stima delle date di separazione dei nodi — beneficia oggi di approcci computazionali integrati che combinano dati linguistici (cognate sets, strutture grammaticali, isoglosse) con dati genomici (frequenze alleliche nelle popolazioni IE moderne e antiche) in un framework bayesiano unificato. L'integrazione e' motivata dalla stretta connessione tra storia linguistica e storia demografica nelle espansioni IE: se una lingua IE si e' diffusa principalmente per movimento di popolazioni (demic diffusion), la sua filogenesi genetica e linguistica dovrebbero essere congruenti; se invece si e' diffusa per acculturazione, le filogenesi possono divergere.

Il framework di inferenza bayesiana integrata e' formalizzato come segue. Sia T_L l'albero filogenetico linguistico con nodi {n_i} e lunghezze di ramo {t_i}, e sia T_G l'albero demografico con nodi {m_j} e lunghezze di ramo {s_j}. Il modello integrato propone che T_L e T_G condividano la stessa topologia (stesso ordine di ramificazione) ma possano avere diverse lunghezze di ramo, con una funzione di penalita' che sanziona le divergenze topologiche:

$$P(\text{data_L}, \text{data_G} \mid T_L, T_G, \text{theta}) = P(\text{data_L} \mid T_L, \text{theta_L}) * P(\text{data_G} \mid T_G, \text{theta_G}) * P(T_L, T_G \mid \text{prior})$$

$$\text{prior} \sim \exp(-\text{lambda} * \text{topological_distance}(T_L, T_G))$$

dove lambda e' un parametro che controlla la forza del vincolo di congruenza. Per lambda = 0, il modello e' completamente separato (lingue e geni evolvono

indipendentemente); per lambda -> inf, il modello richiede congruenza topologica perfetta. La stima ottimale di lambda — ottenuta dalla massimizzazione della verosimiglianza marginale (evidence) del modello integrato — misura il grado di congruenza tra la storia linguistica e la storia demografica nell'espansione IE. I risultati preliminari di questo approccio (Bouchard-Cote et al. 2013 in una versione piu' semplice; Reich et al. 2023 in una versione piu' complessa) suggeriscono un lambda moderato: la filogenesi linguistica e quella genomica sono parzialmente congruenti ma non identiche — coerente con un modello misto in cui la IE-ficazione avviene sia per demic diffusion che per acculturazione in proporzioni variabili per diversi rami.

CAPITOLO XXV — PROSPETTIVE FUTURE: AGENDA DI RICERCA PER IL XXI SECOLO

XXV.1 I problemi aperti fondamentali: una mappa delle incertezze residue

La monografia sulle origini dei Proto-Indoeuropei, nelle sue quattro parti, ha presentato lo stato attuale della ricerca con l'obiettivo di distinguere sistematicamente tra cio' che e' stabilito con alto grado di certezza, cio' che e' probabile ma non definitivo, e cio' che rimane genuinamente aperto. Questo capitolo finale sistematizza le incertezze residue e propone un'agenda di ricerca per il prossimo decennio (2025-2035), identificando i nodi problematici che potrebbero essere risolti con i metodi e le tecnologie attualmente disponibili o di prossima disponibilita'. I problemi aperti fondamentali sono cinque, organizzati per priorita' epistemica.

Problema 1: La risoluzione genomica dell'espansione anatolica. La questione del ramo anatolico — se il proto-anatolico (progenitore di ittita, luvio, palaaico) si e' separato dal tronco IE prima o durante o dopo la formazione dello Yamnaya — e' la piu' acuta delle incertezze residue. I dati aDNA dell'Anatolia del III e II millennio a.C. disponibili (circa 60-80 individui) sono insufficienti per risolverla: le proporzioni di ancestralita' steppa variano e la struttura temporale del campionamento e' disomogenea. La risoluzione richiede un campionamento sistematico dell'Anatolia del Calcolitico finale e del Bronzo Antico (3500-2500 a.C.), con copertura geografica delle aree di potenziale ingresso della componente steppa (Tracia, Gallipoli, Caucaso nord-occidentale).

Problema 2: La risoluzione del proto-greco. Le origini genetiche e linguistiche del proto-greco — la forma di proto-IE che ha dato origine al greco — non sono ancora ricostruite con precisione sufficiente. I dati aDNA della Grecia del Bronzo Medio e Antico (disponibili solo per 30-40 individui) non permettono di stabilire se e quando la componente steppa e' entrata nella penisola ellenica in misura sufficiente a corrispondere all'arrivo del proto-greco. La risoluzione richiede campionamento sistematico della Grecia del III millennio a.C. e dei Balcani settentrionali nello stesso periodo.

Problema 3: Le origini genomiche del proto-italico e del proto-celtico. I rami italico e celtico — che insieme comprendono le lingue piu' ampiamente diffuse della storia mondiale (latino, spagnolo, portoghese, francese, inglese, tedesco) — non hanno una storia genomica chiara nella preistoria. I dati aDNA dell'Italia e della penisola iberica del III-II millennio a.C. sono disponibili ma l'interpretazione in termini di arrivo delle rispettive proto-lingue e' ancora controversa. La connessione tra la cultura Bell Beaker e la proto-celticita' (o la proto-italicita') e' ancora dibattuta.

Problema 4: La struttura interna del proto-IE: dialetti e variazione. Il PIE ricostruito e' tipicamente presentato come un'entita' linguistica omogenea, ma le evidenze di isoglosse che connettono solo alcuni rami IE suggeriscono che gia' all'interno del PBS ci fossero varieta' dialettali. La struttura interna del PIE — i suoi dialetti e le loro distribuzioni geografiche — non e' ricostrubile con i metodi attuali ma potrebbe essere accessibile attraverso l'analisi di reticoli filogenetici (piuttosto che alberi) che permettono di rappresentare le relazioni orizzontali tra i dialetti.

Problema 5: La causalita' tra espansione genomica e espansione linguistica. Il problema fondamentale — che tutti i dati aDNA non risolvono direttamente — e' se la correlazione osservata tra ancestralita' steppa e lingue IE rifletta una causalita' (la presenza di ancestralita' steppa implica diffusione della lingua IE dagli Yamnaya) o una co-associazione (entrambe si diffondono insieme per ragioni diverse). Questa domanda e' strutturalmente analoga al problema della validita' esterna della correlazione genomica-linguistica, e la sua risposta richiede metodi di inferenza causale applicati ai dati aDNA che sono ancora in fase di sviluppo.

XXV.2 Le nuove tecnologie che trasformeranno il campo nel decennio 2025-2035

Il decennio 2025-2035 sara' caratterizzato dall'impatto di tre innovazioni tecnologiche che trasformeranno radicalmente la ricerca sulle origini IE. (1) Il sequenziamento aDNA a ultra-basso costo: la riduzione del costo del sequenziamento genomico, proiettata a continuare lungo la curva di apprendimento esponenziale (legge di Moore genomica), rendera' economicamente fattibile il sequenziamento sistematico di decine di migliaia di individui antichi per regione-periodo. Si stima che entro il 2030 sara' possibile sequenziare a copertura

sufficiente tutti i resti scheletrici disponibili nelle collezioni museali europee e dell'Asia centrale — un corpus di potenzialmente 100.000-200.000 individui antichi. Questo aumento di due ordini di grandezza nella dimensione del dataset risolverebbe molte delle incertezze genomiche attuali, in particolare per le regioni e i periodi attualmente sotto-campionati (Anatolia preistorica, India del II millennio a.C., Asia centrale del III-I millennio a.C.). (2) Il aDNA dai sedimenti (sedaDNA): le tecniche di estrazione di DNA antico dai sedimenti lacustri e dalle grotte, senza richiedere resti scheletrici individuali, permetteranno di ottenere profili genomici di popolazione da qualsiasi sito con buona conservazione tafonomica. Questa tecnica e' gia' operativa per il aDNA di piante e animali (Parducci et al. 2017) e sta mostrando risultati preliminari promettenti per il DNA umano; entro il 2030 potrebbe diventare un metodo routinario che riempira' i gap geografici dove i resti scheletrici sono rari. (3) I modelli linguistici di grande scala (LLM): i modelli di linguaggio neurale di grandissima scala (Large Language Models, LLM) — come GPT-4, Llama, Mistral — addestrati su corpus multilingui che includono testi in lingue antiche, stanno dimostrando capacita' emergenti di analisi morfosintattica e di comparazione lessicale che potrebbero essere sfruttate per la ricostruzione PIE. La fine-tunizzazione di LLM su corpus di linguistica storica IE potrebbe produrre assistenti virtuali per filologi che accelerano significativamente l'analisi di testi nelle lingue IE estinte.

Tecnologia emergente	Applicazione per origini IE	Orizzonte temporale	Impatto atteso	Sfide principali
aDNA ultra-basso costo	Campionamento massiccio resti antichi	2025-2030	Risoluzione questioni genomiche	Accesso collezioni, etica
sedaDNA umano	Profili pop. senza resti scheletrici	2027-2033	Gap geografici colmati	Specificita', contaminazione
LLM per lingue antiche	Assistenza analisi filologica	2024-2028	Produttivita' filologica	Allucinazioni, validazione
Sequenziamento proteomico	Dati su resti dove DNA degradato	2025-2030	Estensione temporale aDNA	Risoluzione tassonomica
Analisi isotopica avanzata	Mobilita' individuale preistorica	2024-2027	Migrazioni individuali	Calibrazione locale

Tecnologia emergente	Applicazione per origini IE	Orizzonte temporale	Impatto atteso	Sfide principali
Reticoli filogenetici linguistici	Struttura dialettale PIE	2025-2030	Dialettologia PIE	Datasets di cognates
ABC con simulazioni massive	Test modelli dispersione IE	2025-2032	Selezione modello rigorosa	Computazione, statistiche sommario
Cryo-EM per DNA degradato	Struttura molecolare DNA antico	2030-2035	Recupero sequenze impossibili	Tecnica ancora sperimentale

XXV.3 Il posto della questione IE nella storia della scienza e nella cultura europea

La questione delle origini indoeuropee non e' un problema esclusivamente tecnico-scientifico: e' anche, e profondamente, un problema culturale e ideologico che ha attraversato la storia europea degli ultimi due secoli con implicazioni che travalicano la linguistica comparata. L'idea di una "razza ariana" che sarebbe all'origine della superiorita' europea — un'idea che non ha nulla a che vedere con le scoperte scientifiche genuine della comparatistica IE, ma che da esse fu politicamente derivata nella Germania del XIX e XX secolo — ha avuto conseguenze tragiche ben note. La storia dell'uso e dell'abuso della questione IE e' essa stessa un oggetto di studio metodologicamente rilevante, perche' illustra come concetti scientifici genuini possano essere deformati per scopi ideologici quando vengono estratti dal loro contesto epistemico.

La scienza contemporanea delle origini IE — basata sulla paleogenomica, sulla linguistica computazionale e sulla filogenesi bayesiana — e' radicalmente diversa nella struttura epistemologica da quella ottocentesca: e' quantitativa, falsificabile, auto-correttiva, e produce risultati probabilistici piuttosto che deterministici. Questa scienza non supporta nessun concetto di "superiorita' razziale" — la mescolanza genomica e' la norma nella preistoria IE come in ogni altra preistoria umana, e le lingue IE si diffondono attraverso processi demografici e culturali che non implicano superiorita' biologica di nessun gruppo. La chiarezza su questo punto non e' solo un obbligo etico ma una conseguenza diretta della corretta applicazione del metodo scientifico: i dati aDNA mostrano che le popolazioni Yamnaya erano esse stesse il

risultato di mescolanza tra EHG e CHG, che le Corded Ware erano mescolanza tra Yamnaya e agricoltori europei, e che ogni popolazione IE storica e' una stratificazione di contributi genomici multipli. La nozione di "purezza etnica" e' semplicemente falsificata dalla genomica.

Il posto della questione IE nella storia della scienza e' quello di un problema genuinamente difficile — abbastanza difficile da resistere a tre secoli di indagine sistematica senza risoluzione definitiva — e abbastanza ricco da generare innovazioni metodologiche importanti in campi diversi. La comparatistica IE ha generato la linguistica storica come disciplina scientifica; la filogenesi IE bayesiana ha contribuito allo sviluppo dei metodi di inferenza filogenetica ora usati in biologia; la paleogenomica IE ha spinto lo sviluppo delle tecniche di sequenziamento aDNA ad alta risoluzione. La questione delle origini IE e', in questo senso, un acceleratore metodologico che beneficia campi ben oltre i propri confini disciplinari.

APPENDICI — PARTE QUARTA

Appendice R — Tavola comparativa sistematica del lessico balto-slavo: selezione per campo semantico

La tabella seguente presenta una selezione sistematica di corrispondenze lessicali tra le lingue baltiche (lituano e lettone) e le principali lingue slave (russo, polacco, ceco, OCS), con la ricostruzione PBS corrispondente e il confronto con il PIE ricostruito. La selezione copre i campi semantici fondamentali per la ricostruzione del vocabolario culturale PBS: natura, corpo umano, parentela, numerali, verbi fondamentali, e lessico culturale.

Significato	PIE ricostruito	PBS ricostruito	Lituano	Lettone	Russo	Polacco	OCS
padre	*pH2ter-	*pater-/*ote-	tevas	tevs	otets	ojciec	otiets
madre	*maHter-	*mater-	motina	mate	mat'	matka	mati
fratello	*bhrH2ter-	*brater-	brolis	bralis	brat	brat	bratu
figlio	*suHnu-	*sunH1u-	sunus	dels	syn	syn	synu
acqua	*wed-or	*wod-a	vanduo	udens	voda	woda	voda
fuoco	*peh2-ur	*paw-ro-	ugnis	uguns	ogon'	ogien	ognj
terra	*dheghom-	*zem-	zeme	zeme	zemlja	ziemia	zemlja
cielo	*dyew-	*dew-	dangus	debess	nebo	niebo	nebo
sole	*seH2wel-	*saule-	saule	saule	solntse	slonce	slunitse
luna	*meh1n-s	*men-	menulis	menuess	mesjats	miesiac	mesjatsu
uno	*H1oi-no-	*edino-	vienas	viens	odin	jeden	edinu
due	*dwoH1	*dwa	du (masc.)	divi	dva	dwa	duva
tre	*treyes	*treje	trys	tris	tri	trzy	trije
dieci	*deHkm	*desmt-	desimit	desmit	desjat'	dziesiec	deseti
cento	*kmtom	*simto-	simtas	simts	sto	sto	suto

Appendice S — Le isoglosse satem: distribuzione sistematica e analisi comparativa

Le lingue "satem" — termine derivato dall'avestico satem "cento" (cf. il latino "centum" del gruppo complementare) — sono quelle lingue IE che hanno palatalizzato le occlusive velari palatali PIE *k', *g', *g'h in spiranti o sibilanti. Il gruppo satem comprende: le lingue baltiche, le lingue slave, le lingue indo-iraniche (sanscrito, avestico, persiano, e le loro discendenti), l'armeno e l'albanese. Il gruppo centum comprende: le lingue anatoliche, il greco, le lingue italiche, le lingue celtiche e le lingue germaniche. Il tocarico, nonostante la sua posizione geografica orientale vicina alle lingue satem, appartiene al gruppo centum, il che costituisce un'evidenza importante per la questione della sua origine: il proto-tocarico deve essersi separato dall'albero IE prima della palatalizzazione satem, o essersi trovato in una posizione geografica che ha evitato la diffusione areale di questo cambiamento.

Esempio	PIE velari palatali	Gruppo satem	Gruppo centum	Distribuzione
cento	*kmtom	Satem: Av. satem, San. satam, Sl. sto	Centum: Lat. centum, Gr. hekaton	Tutte le lingue IE
cuore	*k'erd-	Satem: San. hrd, Sl. serdtse	Centum: Lat. cor, Gr. kardia	Tutte le lingue IE
sapere	*woyd-	Satem: San. veda, Sl. vedeti	Centum: Lat. videre, Gr. oida	Tutte le lingue IE
sorella	*sweg'sor	Satem: San. svasar, Sl. sestra	Centum: Lat. soror, Got. swistar	Tutte le lingue IE
verde	*ghel-	Satem: San. hari-, Sl. zoloto "oro"	Centum: Gr. khloros, Lat. helvus	Tutte le lingue IE
serpente	*H2engwhi-	Satem: Lit. angis, Sl. uzh	Centum: Lat. anguis, OIr. esc	Tutte le lingue IE

Appendice T — Derivazione formale dell'equazione di Fisher-KPP per l'espansione demografica IE

La derivazione dell'equazione di Fisher-Kolmogorov-Petrovskii-Piskunov (Fisher-KPP) applicata all'espansione demografica IE parte dal modello biologico originale di Fisher (1937) per la diffusione di un allele vantaggioso in una

popolazione spazialmente distribuita. Sia u(x,t) la densita' di popolazione con il tratto considerato (nel nostro caso, l'ancestralita' steppa o la proporzione di parlanti IE) nel punto x al tempo t. Il modello combina due processi: diffusione spaziale casuale (movimento di individui nello spazio, modellizzato come diffusione browniana) e crescita logistica (crescita della popolazione con saturazione alla capacita' portante K):

$$\frac{\partial u}{\partial t} = D \cdot \nabla^2 u + r \cdot u \cdot \left(1 - \frac{u}{K}\right) \quad [Fisher - KPP]$$

Per il caso unidimensionale e senza geometria complessa:

du/dt = D * d^2u/dx^2 + r*u*(1 - u/K)

Cerchiamo soluzioni onde viaggianti della forma u(x,t) = phi(x - ct), dove c e' la velocita' dell'onda. Sostituendo nella PDE:

-c * phi' = D * phi'' + r*phi*(1 - phi/K)

con condizioni al bordo phi(-inf) = K (densita' piena di parlanti IE alle spalle del fronte), phi(+inf) = 0 (nessun parlante IE davanti al fronte). L'analisi di stabilita' lineare del punto fisso phi = 0 (fronte dell'onda) mostra che esistono soluzioni onda viaggiante per tutte le velocita' c >= c_min, dove:

$$c_{min} = 2 \cdot \sqrt{D \cdot r}$$

La soluzione selezionata fisicamente (sotto condizioni iniziali con supporto compatto) e' quella con c = c_min — l'onda piu' lenta. Questa e' la velocita' di avanzamento del fronte demografico IE stimata in 1.5-2.5 km/anno per l'espansione Corded Ware. Per calibrare i parametri D e r con i dati aDNA, si minimizza la distanza tra la data radiocarboncia del primo individuo con ancestralita' steppa > 50% in ogni area e la previsione del modello:

$$L(D, r) = \sum_i \frac{\left[t_{model}(x_i; D, r) - t_{obs}(x_i)\right]^2}{\sigma_i^2}$$

(D_hat, r_hat) = argmin_{D>0, r>0} L(D, r) [stima MLE]

I valori ottimali stimati dalla letteratura (Haak et al. 2015; Lazaridis et al. 2016, con calibrazioni su 200+ individui aDNA) sono D ≈ 60-100 km^2/generazione e r ≈ 0.02-0.05/anno, producendo una velocita' di avanzamento c ≈ 2.2-4.5 km/anno —

leggermente superiore alla stima della neolitizzazione (1 km/anno) ma non incompatibile con l'uso del carro e del cavallo come mezzo di trasporto.

Appendice U — Il Corpus dei Testi Sogdiani: selezione e analisi linguistica

Il corpus sogdiano e' distribuito geograficamente lungo la Via della Seta e comprende testi in diversi supporti e generi. La classificazione principale distingue: (1) le Lettere Antiche di Dunhuang (ca. 313-314 d.C.), ritrovate nella Torre del Giada nel 1907 da Aurel Stein; (2) i testi manichei di Turfan (III-VIII secolo d.C.), rinvenuti nelle grotte di Turfan in Xinjiang; (3) i testi buddisti (IV-IX secolo d.C.), trovati principalmente a Dunhuang; (4) i testi cristiani nestoriani (VII-IX secolo d.C.); (5) le iscrizioni rupestri della Via della Seta (Karakoram Highway, I-VIII secolo d.C.); (6) i documenti commerciali e amministrativi dal Monte Mugh in Tagikistan (VIII secolo d.C.), trovati nel 1933 e fondamentali per la comprensione dell'amministrazione sogdiana.

Corpus sogdiano	Data	Localita' ritrovamento	Genere	Pubblicazione principale
Lettere Antiche	~313-314 d.C.	Torre del Giada, Dunhuang	Lettere commerciali privat.	Sims-Williams 1985, 2001
Testi manichei Turfan	~400-900 d.C.	Grotte Turfan, Xinjiang	Testi liturgici manichei	Sundermann 1981-2001
Testi buddisti	~400-900 d.C.	Dunhuang, Xinjiang	Sutra, commentari	Matsuda 1988; Sims-Wil. 2007
Testi nestoriani	~600-900 d.C.	Turfan, Mongolia	Liturgia cristiana	Sims-Williams 1994
Iscrizioni Karakoram	~I-VIII d.C.	Via Seta, Pakistan N.	Graffiti, dediche	Sims-Williams & Cribb 1996
Documenti Monte Mugh	~700-730 d.C.	Tagikistan	Ammin. feudale	Freiman 1962; Bogolyubov 1963

Appendice V — Il Lineare B: sistema di scrittura, segni principali e decodifica di Ventris

La decifrazione del Lineare B da parte di Michael Ventris nel giugno 1952 e' uno degli eventi piu' celebri nella storia della linguistica del Novecento. Ventris — architetto di professione, non filologo accademico — annuncio' in una trasmissione radiofonica della BBC (1 luglio 1952) che il Lineare B era una forma arcaica di greco, confermando un'ipotesi che lui stesso aveva fino a poco prima ritenuto improbabile. La conferma decisiva arrivò dall'analisi della tavoletta PY Ep 704 di Pilo da parte del professore John Chadwick, che identifica la parola ti-ri-po-de (= greco tripodes "treppiedi") seguita da segni numerici e ideogrammici che mostrano effettivamente tre oggetti a tre gambe.

Il sistema del Lineare B ha 87 segni sillabici principali (ciascuno con valore CV o V) e circa 100 ideogrammi (logogrammi per merci, animali, prodotti). Il segno Y45, ad esempio, ha valore "de" e corrisponde alla prima sillaba di termini greci come de-wo-jo (= thewoios "divino"). La struttura sillabica impone le limitazioni fonologiche note: il gruppo consonantico greco ktl- viene scritto come a-to-ro-qo (= anthropos "uomo") con vocali di supporto inter-consonantiche; la consonante finale non e' mai scritta, rendendo il sistema sotto-determinato per la fonologia finale.

Segno LB	Valore fonico	Parola in LB	Parola greca	Significato
a2	a	a-to-ro-qo	anthropos	uomo
e	e	e-re-u-te-re	eleutheroi	liberi
ti	ti	ti-ri-po-de	tripodes	treppiedi
wa	wa	wa-na-ka	wanaks	re, signore
ko	ko	ko-wo	korwos	ragazzo
pa	pa	pa-si	pansi	a tutti
te-o	teo	te-o-i	theoi	dei
po-ti-ni-ja	potnia	po-ti-ni-ja	potnia	signora, dea

Appendice W — Derivazione formale del modello di filogenesi integrata linguistica-genomica

Il framework formale per la stima simultanea degli alberi filogenetici linguistico (T_L) e genomico (T_G) con vincolo di congruenza e' derivato dal framework

bayesiano della phylogenetics integrata. Definizioni formali preliminari. Sia F = {f1, f2, ..., fn} l'insieme delle n lingue IE considerate. Per ogni lingua f_i, il dato linguistico D_L_i e' un vettore binario di cognate presence/absence per M cognate sets: D_L_i in {0,1}^M. Il dato genomico D_G_i e' un vettore di frequenze alleliche per L SNP: D_G_i in [0,1]^L.

Il modello filogenetico linguistico assume evoluzione Stochastic Dollo per i cognates (ogni cognate acquisito una volta, poi perso stocasticamente), con tasso di perdita q_{10} specifico per cognate e orologio molecolare rilassato (tasso di evoluzione diverso per diversi rami dell'albero). La verosimiglianza e' calcolata con l'algoritmo di Felsenstein:

$$L_L\left(T_L, \theta_L | D_L\right) = \prod_{c=1}^{M} P\left(D_L[\cdot, c] | T_L, q_{10,c}, clock_L\right)$$

Il modello filogenetico genomico assume deriva genetica lungo i rami dell'albero, con flusso genico come archi aggiuntivi nel grafo (modello di ammixture). La verosimiglianza genomica e' basata sulle statistiche-f di Patterson:

L_G(T_G, theta_G | D_G) = prod_{j} P(f4(i,j,k,l) | T_G, theta_G)

Il vincolo di congruenza tra T_L e T_G e' espresso come penalita' esponenziale sulla distanza topologica:

P(T_L, T_G) propto exp(-lambda * d_top(T_L, T_G)) * P(T_L) * P(T_G)

dove d_top(T_L, T_G) e' la distanza Robinson-Foulds tra i due alberi (numero di bipartizioni presenti in uno e assente nell'altro, normalizzata per 2(n-3)), e P(T_L) e P(T_G) sono i prior separati sulle topologie (tipicamente prior Yule per T_L e prior uniforme o coalescente per T_G). La distribuzione posteriore congiunta e':

P(T_L, T_G, theta | D) propto L_L * L_G * P(T_L, T_G) * P(theta)

Il campionamento dalla distribuzione posteriore congiunta e' effettuato con MCMC che alterna mosse su T_L (SPR - Subtree Prune and Regraft), mosse su T_G (simili ma con gestione degli archi di ammixture), e mosse sui parametri theta (Metropolis-Hastings standard). La diagnostica di convergenza usa il criterio di Gelman-Rubin su catene multiple indipendenti (R-hat < 1.01 per tutti i parametri). Il calcolo del posteriore marginale su lambda (il parametro di congruenza) fornisce la

stima del grado di congruenza tra filogenesi linguistica e genomica, che e' il risultato di interesse principale.

Appendice X — Tavola sinottica degli script IE: comparazione dei sistemi di scrittura

Sistema	Tipo	Dir. scrittura	N. segni	Lingue	Periodo uso	Status
Cuneiforme ittita	Logo-sillabico	Sin. > Des.	~375	Ittita, Luvio, Palaaico	1700-1200 a.C.	Estinto
Lineare B	Sillabico	Sin. > Des.	~90 segni + ideo.	Greco miceneo	1450-1200 a.C.	Estinto
Geroglifico luvio	Logo-sillabico	Bustrofedico	~500	Luvio Neo-ittita	1200-700 a.C.	Estinto
Alfabeto greco	Alfabeto vocalico	Sin. > Des.	24	Greco e derivati	~800 a.C. - oggi	Vivo
Alfabeto latino	Alfabeto vocalico	Sin. > Des.	23-26	Latino e tutti occ.	~600 a.C. - oggi	Vivo
Devanagari	Abugida	Sin. > Des.	~50	Hindi, Sanscr., Marathi	~1000 d.C. - oggi	Vivo
Futhark runico	Alfabeto	Variabile	24-33	Germ. settentrionali	~150-1200 d.C.	Storico
Cirillico	Alfabeto vocalico	Sin. > Des.	33 (russo)	Slavo orientale/merid.	~863 d.C. - oggi	Vivo
Script sogdiano	Abjad adattato	Des. > Sin.	~22	Sogdiano	~200-1000 d.C.	Estinto/derivato
Avestico (Pahlavi)	Alfabeto modificato	Des. > Sin.	~53	Avestico	~400-900 d.C.	Liturgico

Appendice Y — Glossario avanzato dei termini tecnici della Parte Quarta

Il seguente glossario completa i glossari delle parti precedenti con i termini tecnici introdotti nella Parte Quarta, organizzati per campo disciplinare.

Linguistica storica e comparativa:

ABLAUT (o Apofonia): alternanza vocalica sistematica all'interno di paradigmi morfologici, tipicamente tra grado e (*-e-), grado o (*-o-), e grado zero (*-Ø-), correlata con la posizione dell'accento nel PIE.

ABJAD: sistema di scrittura in cui ogni segno rappresenta una consonante e le vocali sono omesse o indicate con segni diacritici supplementari; tipico delle scritture semitiche (aramaico, ebraico, arabo).

ABUGIDA: sistema di scrittura in cui ogni segno rappresenta una consonante con una vocale di default incorporata, e le altre vocali sono indicate con modificazioni del segno base; tipico della Devanagari e delle scritture indiane.

BUSTROFEDICO: sistema di scrittura in cui le linee alternano direzione (da sinistra a destra poi da destra a sinistra, come i buoi che arano un campo), usato in alcune iscrizioni greche arcaiche e nei geroglifici luvi.

ISOGLOSSA: linea immaginaria che demarca l'area di diffusione di un fenomeno linguistico specifico (fonologico, morfologico o lessicale) in una mappa geografica; le convergenze di isoglosse definiscono i raggruppamenti dialettali.

METATESI: trasposizione dell'ordine di due segmenti adiacenti in una parola; la metatesi dei liquidi (liquide + vocale > vocale + liquide) e' tipica del proto-slavo (*TorT > *ToroT).

OPACO/TRASPARENTE (morfologia): una forma morfologicamente opaca non e' piu' analizzabile in morfemi riconoscibili; una forma trasparente mostra la sua struttura morfologica in modo evidente. L'erosione morfologica e' il processo per cui forme trasparenti diventano opache nel corso del tempo.

POSIZIONE WACKERNAGEL: posizione di seconda posizione nella frase (dopo il primo costituente fonologico), occupata sistematicamente dai clitici in molte lingue IE arcaiche, in accordo con la legge di Wackernagel (1892).

Genomica e paleogenomica:

ANE (Ancient North Eurasian): componente ancestrale genomica documentata in cacciatori-raccoglitori siberiani del Paleolitico superiore (Mal'ta, 24.000 a.C.), distinta da EHG e WHG; particolarmente alta nelle mummie del Tarim.

SEDADNA (Sedimentary Ancient DNA): DNA antico estratto da sedimenti lacustri o da depositi di grotte, senza richiedere resti scheletrici individuali; permette di ottenere profili genomici di popolazione per siti dove i resti ossei sono assenti o mal conservati.

PALAEOPROTEOMICA (o Proteomica Antica): analisi delle proteine conservate in resti biologici antichi, compresa la dentina ossea; piu' stabile del DNA nel tempo e quindi applicabile a campioni piu' antichi (potenzialmente milioni di anni).

Linguistica computazionale:

EXACT MATCH (accuratezza): metrica di valutazione di sistemi NLP che conta la percentuale di predizioni identiche al gold standard carattere per carattere (o token per token); metrica severa perche' non riconosce varianti quasi-corrette.

EDIT DISTANCE (distanza di Levenshtein): numero minimo di operazioni di inserzione, cancellazione e sostituzione di caratteri necessario per trasformare una stringa nell'altra; usata come metrica soft di accuratezza nella ricostruzione fonologica.

FINE-TUNING: procedura di addestramento aggiuntivo di un modello di linguaggio neurale pre-addestrato su un corpus specifico di dominio (ad esempio testi di linguistica storica IE), permettendo al modello di acquisire conoscenza specializzata mantenendo le capacita' generali del pre-addestramento.

GOLD STANDARD: nel contesto della valutazione di sistemi NLP, l'insieme di output ritenuti corretti per un dato task, tipicamente prodotti da esperti umani e usati come riferimento per misurare l'accuratezza del sistema automatico.

Appendice Z — Cronologia sinottica aggiornata e bibliografia critica della Parte Quarta

La tavola cronologica seguente integra le tavole delle Parti precedenti con gli eventi rilevanti per i capitoli della Parte Quarta, con enfasi su: (1) storia delle lingue balto-slave; (2) collasso del Bronzo Tardo; (3) lingue IE dell'Asia centrale; (4) sviluppo dei sistemi di scrittura IE; (5) sviluppi computazionali recenti.

Data (ca.)	Evento	Rilevanza per le origini IE
3000-2500 a.C.	Espansione Corded Ware in Europa centrale e baltica	Arrivo PBS nell'area baltica; inizio IE-ficazione nord-europea
2500-2000 a.C.	Diversificazione Proto-Balto-Slavo	Separazione protoceltica/protoitalica dalle culture Bell Beaker
2200-1900 a.C.	Evento climatico 4.2 ka: siccita' globale	Possibile accelerazione migrazioni IE; collasso civilta' antiche
2100-1500 a.C.	Cultura Sintashta-Andronovo; proto-indo-iranico	Espansione indo-ariana; antenati vedici
1700-1200 a.C.	Impero Ittita; Lineare B miceneo; Mitanni	Apogeo lingue anatoliche IE; primo greco scritto; indo-ariano Mitanni
1200-1150 a.C.	Collasso del Bronzo Tardo; fine Yamnaya tardo	Scomparsa ittita, miceneo; popoli del mare; redistribuzione IE
1000-500 a.C.	Cultura La Tene; espansione celtica; formazione etnico-IE	IE-ficazione Gallia, Iberia, Isole Britanniche
900-500 a.C.	Lingue iraniche nelle steppe (Sciti, Sarmati)	Iranizzazione Asia centrale; predecessori sogdiano/battriano
800-400 a.C.	Alfabeto greco; Omero scritto; frigio e trace attestati	Inizio documentazione alfabetica IE; lingue balcaniche minori
600-100 a.C.	Espansione romana; latino come lingua franca	IE-ficazione mediterranea occidentale per via latina
~500 a.C. - 300 d.C.	Tocarico in Bacino Tarim; testi sogdiani Via Seta	IE nell'Asia centrale orientale; lingua franca commerciale
~300-900 d.C.	Espansione slava; cirillico (863 d.C.)	IE-ficazione Europa orientale e meridionale (Balcani)
~1000-1300 d.C.	Documentazione lingue baltiche; diffusione slavo	Primo lituano scritto; diversificazione lingue slave
1786	W. Jones: parentela IE in allocuzione Asiatic Society	Fondazione linguistica comparata IE come disciplina
1819	Franz Bopp: sistema coniugazione sanscrito e lingue IE	Prima grammatica comparata sistematica IE

Data (ca.)	Evento	Rilevanza per le origini IE
1822	Jacob Grimm: prima formulazione legge consonantica	Metodo comparativo applicato al germanico
1876	Scoperta legge di Verner; neogrammatica a Lipsia	Sistematizzazione leggi fonetiche senza eccezioni
1896	Saussure: corrispondenza balto-slava e PIE	Accentologia PIE; laringali PIE (predetti)
1915	Hrozny: decifrazione dell'ittita come lingua IE	Scoperta ramo anatolico IE; confutazione modelli mono-lineari
1952	Ventris: decifrazione Lineare B come greco miceneo	Rivelazione del greco miceneo; ampliamento corpus IE
1956-1970	Gimbutas: ipotesi kurganica sistematizzata	Urheimat steppe formulata; modello dominante
1987	Renfrew: ipotesi anatolica	Sfida alla kurganica; neolitizzazione come vettore IE
2015	Haak et al., Allentoft et al.: aDNA rivoluzione steppe	Conferma genomica espansione Yamnaya-Corded Ware
2021-2023	Heggarty et al. 2023; Anthony et al. 2022; Zhang 2021	Nuovi dati genomici mummie Tarim; filogenesi Bayesian aggiornata
2024-2030	Sequenziamento massivo aDNA; LLM per lingue antiche	Prospettive di risoluzione incertezze genomiche e filologiche

La bibliografia critica seguente integra le opere citate nelle parti precedenti con i lavori fondamentali per i capitoli specifici della Parte Quarta.

Per il ramo balto-slavo (Capitolo XVI):

Kortlandt, F. (1979). "Toward a reconstruction of the Balto-Slavic verbal system." Lingua 49: 51-70. — Dijk, T. van & Kortlandt, F. (2012). The origin of Slavic palatalization. Amsterdam. — Fortson IV, B. (2004). "Indo-European Language and Culture." Blackwell, cap. 12-13. — Meillet, A. (1925). "Le Slave commun." Paris, Champion. — Shevelov, G. (1965). "A Prehistory of Slavic." Heidelberg: Carl Winter.

Per la sintassi PIE (Capitolo XVII):

Lehmann, W.P. (1974). "Proto-Indo-European Syntax." University of Texas Press. — Watkins, C. (1976). "Towards Proto-Indo-European Syntax." Papers from the

Parasession on Diachronic Syntax. Chicago: CLS. — Wackernagel, J. (1892). "Uber ein Gesetz der indogermanischen Wortstellung." IF 1: 333-436. — West, M.L. (2007). "Indo-European Poetry and Myth." Oxford University Press.

Per i contatti linguistici (Capitolo XVIII):

Hyllested, A. (2010). "Internal reconstruction vs. external evidence." Proceedings of the 21st Annual UCLA IE Conference. — Parpola, A. (1994). "Deciphering the Indus Script." Cambridge University Press. — Bomhard, A. (2008). "Reconstructing Proto-Nostratic." Leiden: Brill.

Per il collasso del Bronzo e le scritture (Capitoli XIX-XXII):

Cline, E. (2014). "1177 B.C.: The Year Civilization Collapsed." Princeton University Press. — Ventris, M. & Chadwick, J. (1973). "Documents in Mycenaean Greek." 2nd ed. Cambridge. — Hooker, J. (1980). "Linear B: An Introduction." Bristol Classical Press. — Powell, B. (1991). "Homer and the Origin of the Greek Alphabet." Cambridge.

Per il Sogdiano e l'Asia Centrale (Capitolo XXIII):

Sims-Williams, N. (2000-2001). "Bactrian Documents from Northern Afghanistan." 2 vols. London: Nour Foundation. — Zhang, F. et al. (2021). "Ancient DNA from Tarim Basin." Nature 599: 256-261. — Lurje, P. (2010). "Personal Names in Sogdian Texts." Vienna.

Per la linguistica computazionale (Capitolo XXIV):

Meloni, C. et al. (2021). "Ab Antiquo: Neural Proto-Language Reconstruction." Proceedings of ACL 2021. — Assael, Y. et al. (2022). "Restoring and attributing ancient texts using deep neural networks." Nature 603: 280-283. — Ciobanu, A. & Dinu, L. (2014). "Automatic detection and production of related words for historical linguistics." COLING 2014.

CAPITOLO XXVI — IL RAMO ANATOLICO IN DETTAGLIO: ITTITA, LUVIO, PALAITICO E LICIO

XXVI.1 Il ramo anatolico: posizione filogenetica e importanza per la ricostruzione del PIE

Il ramo anatolico occupa nella famiglia indoeuropea una posizione che nessun altro gruppo linguistico puo' eguagliare per centralita' epistemica: le lingue anatoliche — ittita, luvio, palaitico, liwio geroglifico, lidio, licio, caro, sideto — conservano strutture morfologiche e fonologiche che permettono di risalire a stadi del Proto-Indoeuropeo anteriori a quelli documentabili attraverso qualsiasi altro ramo. La loro separazione precoce dalla famiglia comune, stimata intorno al 4000-3500 a.C. secondo le analisi bayesiane piu' recenti, ha preservato tratti arcaici che le lingue piu' recenti hanno perso attraverso innovazioni condivise (isoglosse) che definiscono i raggruppamenti interni alla famiglia.

La scoperta delle lingue anatoliche e' uno degli episodi piu' drammatici nella storia della linguistica comparata. Le prime tavolette ittite vennero portate sul mercato antiquario intorno al 1880, ma la loro decifrazione avvenne solo nel 1915, quando il ceco Bedrich Hrozny pubblico' la dimostrazione che la lingua delle tavolette di Bogazkoy (l'antica Hattusa, capitale dell'impero ittita) era imparentata con le lingue indoeuropee note. La dimostrazione di Hrozny si baso' su un passaggio diagnostico: la frase bettli-ma adanna harmi («adesso mangio pane») conteneva la radice *h2ed- «mangiare» e il termine per «acqua» (wadar, comparabile con il tedesco Wasser, il greco hudor, il gotico wato) in costruzione con la congiunzione nu/ma, parallela a quella delle lingue IE classiche. La verifica successiva da parte di Ferdinand Sommer e poi di Georg Forrer nei decenni 1920-1940 stabili' il corpus fondamentale della grammatica ittita e la sua posizione nella famiglia IE.

XXVI.1.1 Il problema del ramo anatolico come IE arcaico versus IE innovato: la disputa Sturtevant-Kuryłowicz

La scoperta delle lingue anatoliche genero' immediatamente un dibattito fondamentale che e' tutt'ora irrisolto nella sua portata teorica: le lingue anatoliche sono un ramo dell'albero IE come gli altri rami, oppure rappresentano la derivazione piu' diretta da uno stadio anteriore da cui tutto il resto della famiglia si e' separato

insieme come un'unica entita'? Il filologo americano Edgar Howard Sturtevant propugno' nel 1926 l'ipotesi della «Indo-Hittite» hypothesis: l'anatolico e' il ramo sorella di tutte le altre lingue IE considerate insieme (il cosiddetto «Common IE» o «Proto-IE proper»), e quindi la radice dell'albero cade nel nodo che separa l'anatolico dal resto. Questa ipotesi implica che molte caratteristiche che si pensava fossero innovazioni condivise di tutta la famiglia siano in realta' innovazioni del Common IE, assenti nell'anatolico perche' queste lingue si erano gia' separate.

La posizione opposta — sostenuta da Kuryłowicz, poi da Calvert Watkins e dalla tradizione americana dominante — e' che l'anatolico e' semplicemente un ramo come gli altri, sia pure arcaico, e che l'albero IE non ha una divisione binaria primaria tra anatolico e tutto-il-resto. Le analisi bayesiane contemporanee, in particolare quella di Bouckaert et al. (2012) e di Heggarty et al. (2023), producono risultati che supportano piu' chiaramente la posizione «Indo-Hittite»: in quasi tutti gli alberi campionati dal MCMC, l'anatolico occupa la posizione basale come outgroup del resto della famiglia. Tuttavia l'incertezza residua e' ancora significativa, e la questione non puo' essere considerata definitivamente chiusa.

Le implicazioni di questa disputa sono profonde. Se il PIE «standard» ricostruito dalla tradizione comparatistica ottocentesca (con le laringali, il sistema nominale a otto casi, la morfologia verbale elaborata) e' in realta' il Common IE e non il PIE reale (che includeva l'anatolico), allora il PIE reale era piu' semplice morfologicamente. L'ittita non ha il genere femminile (solo comune/animato e neutro), non ha il subjunctivo, non ha i tempi composti, non ha la distinzione perfetto/imperfetto: tutte queste categorie sarebbero state innovazioni del Common IE posteriori alla separazione anatolica, non tratti del PIE originale.

Categoria	PIE ricostruito standard	Ittita attestato	Status
Genere	3 generi: masc./femm./neutro	Solo: comune / neutro	Femminile: innovazione IE non-anatolica
Casi nominali	8 casi (nom., gen., dat., acc., voc., loc., strum., abl.)	7 casi (mancante il locativo distinto)	Parzialmente conservativo
Modo subjuntivo	Presente, aoristo, perfetto	Assente	Innovazione Common IE

Categoria	PIE ricostruito standard	Ittita attestato	Status
Aspetto verbale	3 aspetti: perfettivo/imperfettivo/stativo	Sistema binario semplificato	Conservativo (sistema piu' antico?)
Laringali	H1, H2, H3 (per ricostruzione)	H2 e H3 attestate come h/hh	Prova diretta della teoria laringale
Participio	Sistema ricco	Partecipale -ant- conservato	Arcaico
Raddoppiamento	Perfetto e intensivo	Solo alcune forme	Innovazione Common IE
Pronomi	Sistema PIE quasi completo	Sistema semplificato ma comparabile	Medio

XXVI.2 La grammatica ittita: morfologia nominale e verbale in dettaglio

XXVI.2.1 Il sistema nominale ittita: declinazione e genere

Il sistema nominale ittita e' organizzato attorno a due generi grammaticali — il genere comune (animato) e il genere neutro — e a sette o otto casi, a seconda dell'analisi. Questa struttura e' significativamente piu' semplice del sistema a tre generi e otto casi ricostruito per il PIE standard, ma conserva tratti fonologici di grande arcaismo. La declinazione del sostantivo ittita si divide in due classi principali: i temi consonantici (il tipo piu' arcaico, corrispondente ai temi in consonante del PIE) e i temi vocalici in -a e in -i/-u.

Prendiamo come esempio il paradigma del sostantivo ishaz- («bocca», tema in -az-), che mostra la struttura di base della declinazione:

Caso	Singolare	Plurale	Comparazione PIE
Nominativo comune	isha-s	isha-s / ish-e-s	PIE *os (nom. sg. animato)
Accusativo comune	isha-n	isha-us	PIE *om/*oms
Genitivo	isha-sh	isha-an	PIE *os (gen. sg.)

Caso	Singolare	Plurale	Comparazione PIE
Dativo/Locativo	isha-i	isha-as	Sincretismo PIE dat.+loc.
Ablativo	isha-az / -anza	—	PIE abl. *-od
Vocativo	isha / isha-e	—	PIE voc. sg. (tonica)
Strumentale	isha-it	isha-it	PIE strum. *-e/o-bhi
Nom./Acc. neutro	isha-n	isha	PIE nom./acc. n. *-om/*-a

Il tratto piu' notevole di questo sistema e' la preservazione dell'ablativo in -az/-anza, che ha un corrispondente diretto nel PIE *-od (attestato anche nel latino arcaico: in -od in formule prelegali come «sei tu cittadino romano? sceis me tuom esse clientem?»). L'ablativo e' perduto nella maggior parte delle lingue IE moderne ma e' conservato nell'ittita in forma quasi inalterata, confermando la sua presenza nel PIE.

XXVI.2.2 Il verbo ittita: il sistema mi-/hi- e la sua interpretazione

La caratteristica piu' distintiva e teoricamente problematica del verbo ittita e' la divisione delle coniugazioni verbali in due classi: il paradigma mi- (dove la prima persona singolare del presente indica finisce in -mi) e il paradigma hi- (dove la prima persona singolare del presente indica finisce in -hi o -hhi). Questa divisione non ha un parallelo diretto nelle altre lingue IE classiche, e la sua spiegazione e' stata oggetto di controversia per quasi un secolo di ittitologia.

La proposta piu' accreditata, formulata da Erich Neu e sviluppata da J. Jasanoff nel suo studio monumentale del 2003, e' che il paradigma hi- riflette direttamente il perfetto PIE. Il perfetto PIE era una categoria aspettuale (piuttosto che temporale) che esprimeva uno stato presente risultante da un'azione passata — esattamente il senso di molti verbi hi- in ittita (molti verbi di stato: «essere seduto», «stare in piedi», «sapere»). In greco, il perfetto aveva sviluppato un valore stativo (oida, «so», e' tecnicamente un perfetto «io ho visto [quindi so]»), in latino il perfetto e' confluito con l'aoristo, in sanscrito il perfetto aveva funzione narrativa. In ittita, questa categoria si e' preservata come classe morfologica produttiva, distinta dal presente-futuro mi-.

PARADIGMA mi- (es. ep- "prendere"):

1sg: e-ep-mi 1pl: e-ep-we-ni

2sg: e-ep-si 2pl: e-ep-te-ni

3sg: e-ep-zi 3pl: a-ap-pa-an-zi

PARADIGMA hi- (es. daa- "prendere/tenere", stativo):

1sg: da-a-hhi 1pl: ta-a-hu-we-ni

2sg: ta-a-at-ti 2pl: ta-a-at-te-ni

3sg: da-a-i 3pl: ta-a-an-zi

La distribuzione dei verbi tra i due paradigmi e' parzialmente lessicale (verbi di stato tendono ad essere hi-) ma anche morfologicamente determinata (derivati con certi suffissi appartengono sistematicamente a una classe o all'altra). La spiegazione diacronica di Jasanoff colloca l'origine della distinzione nel PIE, dove i verbi stativi/risultativi formavano il perfetto con terminazioni -h2a (1sg), -th2e (2sg), -e (3sg) — terminazioni che diventano -hhi, -tti, -i in ittita regolarmente secondo le leggi fonetiche anatoliche.

XXVI.3 Il luvio e le lingue luwie: geroglifico e cuneiforme

Il luvio — conosciuto in due varieta' grafiche, il luvio cuneiforme (scritto con il sistema cuneiforme babilonese/assiro) e il luvio geroglifico (scritto con un sistema ideografico-sillabico sviluppato nella regione) — e' la lingua anatolica con la distribuzione geografica piu' ampia e probabilmente con il maggior numero di parlanti nell'eta' del Bronzo tardo. Il luvio geroglifico e' attestato su monumenti rupestri, stele e sigilli dall'Asia Minore centrale e meridionale fino alla Siria settentrionale, e la sua produzione testuale continua nell'eta' del Ferro anche dopo il crollo dell'impero ittita (~1180 a.C.), attraverso gli stati neo-ittiti della Siria settentrionale (che erano politicamente e culturalmente luwie, non ittiti).

La posizione del luvio rispetto all'ittita e' di lingua sorella, non di dialetto o discendente. Le due lingue si separarono probabilmente intorno al 2500-2200 a.C. (su base cronologica interna, dalla comparazione delle innovazioni condivise) e mostrano differenze significative a tutti i livelli: fonologico (il luvio ha sviluppato un'affricata */ts/ da certi gruppi consonantici IE), morfologico (il sistema dei casi e'

piu' ridotto del ittita) e lessicale (il luvio ha un vocabolario parzialmente diverso, con molti prestiti da lingue vicino-orientali non-IE).

Il luvio e' di grande importanza per la questione del substrato pre-greco. Molte delle parole greche con suffisso -nth-, -ss-/-tt- (come nel greco Korinthos, Athenai, thalassa, elaion, keramikos) — a lungo attribuite a un misterioso «substrato pelasgico» non identificato — corrispondono a modelli morfologici luwie. Questa osservazione, sviluppata da Melchert, Hajnal e Watkins negli anni 2000, suggerisce che il substrato pre-greco in parte possa essere luvio (o una lingua strettamente imparentata col luvio), portato dai movimenti di popolazione del tardo Bronzo nel Mediterraneo orientale.

Lingua	Periodo attestazione	Script	Area geografica	Corpus approx.	Caratteristica chiave
Ittita	1650-1180 a.C.	Cuneiforme babilonese	Anatolia centrale (Hattusa)	~30.000 tavolette	Lingua ufficiale dell'impero ittita
Luvio cuneiforme	1650-1180 a.C.	Cuneiforme	Anatolia merid./occid.	~200 testi	Lingua rituale ittita
Luvio geroglifico	1400-700 a.C.	Geroglifico anatolico	Anatolia/Siria nord.	~700 iscrizioni	Continua dopo crollo ittita
Palaitico	1650-1400 a.C.	Cuneiforme	Paphlagonia (nord Anatolia)	~30 testi brevi	Il piu' arcaico ramo anatolico noto
Licio	500-300 a.C.	Alfabeto derivato dal greco	Licia (Anatolia SW)	~170 iscrizioni	Lingua di substrato del greco
Lidio	700-200 a.C.	Alfabeto lidio	Lidia (Anatolia occ.)	~100 iscrizioni	Alta enigmaticita' lessicale
Cario	600-200 a.C.	Alfabeto cario (decifrato 1993)	Caria (Anatolia SW)	~200 brevi iscr.	Decifrazione tardiva e parziale
Sideto	200 a.C.-200 d.C.	Alfabeto locale	Side (Panfilia)	~20 iscrizioni	Posizione incerta nella famiglia

XXVI.4 Il licio: struttura morfologica e posizione nell'albero anatolico

Il licio e' attestato in circa 170 iscrizioni, prevalentemente funerarie, distribuite nel territorio dell'odierna provincia turca di Antalya (l'antica Licia) e databili tra il V e il III secolo a.C. La lingua e' ben decifrata grazie all'ausilio di alcune iscrizioni bilingui (licio-greco, licio-aramaico) e alla sua evidente appartenenza alla famiglia anatolica, che permette di ricorrere ai paralleli ittiti e luwei per identificare morfemi e parole.

Il licio si distingue dalle altre lingue anatoliche per una serie di innovazioni fonologiche notevoli: la perdita generalizzata delle occlusive sonore ereditate (*b, *d, *g diventano rispettivamente p/b, t/d, k/g con neutralizzazione parziale), lo sviluppo di una serie di nasali sillabiche (m, n) da gruppi consonantici PIE, e un sistema vocalico che ha subito numerose sincopi (cadute di vocali interne). Dal punto di vista morfologico, il licio ha mantenuto la distinzione comune/neutro come l'ittita, ma ha semplificato ulteriormente il sistema dei casi a cinque o sei forme, alcune delle quali sono sincretismi di categorie precedenti.

La relazione del licio con il luvio e' strettamente parallela a quella del portoghese con lo spagnolo: le due lingue sono chiaramente imparentate, con numerosi cognati (il termine licio se- «e', egli» corrisponde al luvio a-sa-; il licio hri «per, su» corrisponde al luvio ari «su»), ma con innovazioni indipendenti sufficienti per costituire rami distinti. L'analisi cladistica bayesiana colloca il licio come ramo sorella del luvio (insieme all'iscrizio del Milya, o miliaco, spesso considerato un dialetto licio), con il palaitico come ramo piu' basale all'interno del clade anatolico non-ittita.

Un tratto di grande interesse del licio e' la sua conservazione di arcaismi PIE assenti nell'ittita. In particolare, il licio documenta il suffisso nominale *-nt- in forme come mahanahi «re, signore» (tema in *-nt-) e il suffisso verbale dell'imperfetto in forma conservativa. Questi arcaismi, combinati con le innovazioni proprie del licio, producono il paradosso apparente di una lingua che e' al tempo stesso piu' arcaica e piu' innovata rispetto all'ittita — un paradosso che si risolve considerando che le diverse strutture si trovano a livelli diversi della grammatica.

CAPITOLO XXVII — IL MITANNI-ARIANO E LE LINGUE IE NEL VICINO ORIENTE ANTICO

XXVII.1 La documentazione del Mitanni-ariano: un IE arcaico in Mesopotamia

La scoperta di un substrato linguistico indo-ariano nel Vicino Oriente del II millennio a.C. e' uno degli sviluppi piu' sorprendenti e informativi della comparatistica indoeuropea del XX secolo. Le tavolette di Amarna (1887), gli archivi di Bogazkoy (scavati dal 1905) e soprattutto le tavolette di Nuzi (pubblicate negli anni 1930-1940) hanno rivelato l'esistenza di un'elite di cavalieri e allevatori di cavalli — i Maryannu — che usava parole tecniche e nomi propri di evidente origine indo-ariana in un contesto linguistico altrimenti hurrico (nel caso del regno di Mitanni) o babilonese (nelle regioni limitrofe).

Il corpus del Mitanni-ariano — termine introdotto da Paul Thieme nel 1960 per designare l'elemento indo-ariano attestato nei testi hurrici del Mitanni — comprende tre categorie principali di evidenza. La prima e' quella dei teonimi: un trattato di pace tra il re mitannico Mattiwaza e il re ittita Suppiluliuma (datato ~1380 a.C.) invoca come garanti divini Mitra, Varuna, Indra e Nasatya — le quattro divinita' cardinali del pantheon vedico, trascritte in forma cuneiforme come Mi-it-ra, U-ru-wa-na, In-da-ra, Na-sa-at-ti-ia-an-na. Questo testo e' la prova piu' diretta e incontrovertibile della presenza di una tradizione religiosa proto-vedica nell'area siro-mesopotamica.

La seconda categoria e' quella dei termini tecnici relativi all'addestramento dei cavalli da guerra. Un trattato sull'addestramento degli stalloni, scritto in hurrico ma con numerosi termini tecnici indo-ariani traslitterati, e' attribuito a un certo Kikkuli (probabilmente un nome hurricizzato) ed e' datato al XIV secolo a.C. Il trattato di Kikkuli usa termini come aika- (cf. vedico eka-, «uno»), tera- (cf. vedico tri-, «tre»), panza- (cf. vedico panca, «cinque»), satta- (cf. vedico sapta, «sette»), na-(va)-wartanna- (cf. vedico nava-vartana «nove giri [di pista]») per indicare le sessioni di esercizio dei cavalli — un vocabolario tecnico numerale che dimostra la connessione con il ramo indo-ariano specificamente (non con il ramo iranico, che ha forme diverse per gli stessi numeri).

Termine Kikkuli	Traslitterazione	Vedico corrispondente	Significato	Commento
aika-wartanna	ai-ik-ka-war-tan-na	eka-vartana	Un giro (di pista)	eka < PIE *oino- «uno»
tera-wartanna	te-ra-war-tan-na	tri-vartana	Tre giri	tri < PIE *treyes
panza-wartanna	pan-za-war-ta n-na	panca-vartana	Cinque giri	panca < PIE *penkWe
satta-wartanna	sat-ta-war-tan -na	sapta-vartana	Sette giri	sapta < PIE *septm
na-wartanna	na-a-war-tan-na	nava-vartana	Nove giri	nava < PIE *H1newn
Maryannu	Ma-ar-ia-an-n u	marya (vedico)	Guerriero/uomo giovane	marya < PIE *moryos
Mitra (teonimo)	Mi-it-ra	Mitra	Dio del contratto	mitra < PIE *mei- «legare»
Indra (teonimo)	In-da-ra	Indra	Dio della guerra/pioggia	Etimologia PIE controversa
Varuna (teonimo)	U-ru-wa-na	Varuna	Dio dell'ordine cosmico	Cf. greco Ouranos
Nasatya (teonimo)	Na-sa-at-ti-ia	Nasatya (Ashvin)	Gli Dei Cavalieri Gemelli	na-satya «senza falso»?

XXVII.2 Implicazioni del Mitanni-ariano per la cronologia e la geografia del PIE

L'evidenza del Mitanni-ariano ha implicazioni profonde per la ricostruzione della cronologia e della geopolitica delle migrazioni IE. Il fatto che una componente indo-ariana fosse presente in Siria e Mesopotamia nel XVI-XIV secolo a.C. — in posizione di elite guerriera specializzata nella gestione di cavalli da guerra — richiede la postulazione di una diaspora indo-ariana in direzione ovest e sud rispetto al presumibile punto d'origine nell'area delle steppe euro-asiatiche, parallela o anticipatrice rispetto alla diaspora verso il subcontinente indiano attestata dal Rigveda.

Questa evidenza e' compatibile con due scenari alternativi. Nel primo scenario — quello tradizionale e tuttora dominante — i proto-indo-ariani si espansero dalle steppe dell'Asia centrale (cultura Sintashta-Andronovo) in multiple direzioni contemporaneamente: verso l'India a sud-est, verso il Vicino Oriente a ovest (il clade Mitanni), e verso l'Iran a sud (i proto-iranici). La biforcazione indo-ariana produsse quindi almeno due sotto-rami: uno che si mosse verso l'India (proto-vedico) e uno che si mosse verso la Mesopotamia (Mitanni-ariano), separatisi dalla radice comune prima del XVI secolo a.C.

Nel secondo scenario — proposto da Anthony (2007) e discusso criticamente da Witzel (2003) e da Parpola (2015) — il Mitanni-ariano rappresenta una migrazione piu' precoce dalla regione della cultura BMAC (Bactria-Margiana Archaeological Complex, ~2300-1700 a.C.) verso ovest, mentre la migrazione verso l'India e' posteriore. In questo schema, la stratificazione lessicale del Rigveda (con prestiti da lingue BMAC identificati da Witzel) testimonia un'interazione piu' duratura tra proto-indo-ariani e le popolazioni agricole dell'Asia centrale, mentre il ramo mitannico si separo' prima di questa interazione.

Dal punto di vista linguistico, il Mitanni-ariano mostra tratti piu' arcaici del vedico su certi punti (la conservazione di *a dove il vedico ha subito la biforcazione *a/*i/*u dal PIE), ma piu' innovati su altri (la perdita di certi morfemi verbali). Questa combinazione di tratti arcaici e innovati e' tipica di un ramo che si e' separato precocemente dal tronco comune e poi si e' sviluppato in isolamento. La posizione filogenetica del Mitanni-ariano nell'albero indo-ariano e' probabilmente quella di un ramo basale separato prima o durante la fase del sanscrito antico, ma le dimensioni ridotte del corpus rendono qualsiasi analisi cladistica formale insufficiente.

XXVII.3 L'hurrico e il Mitanni: la lingua di substrato e la struttura sociolinguistica

Il Mitanni-ariano non e' la lingua del regno di Mitanni nel suo insieme: e' un substrato o superstrato utilizzato da una elite guerriera (i Maryannu, i cavalieri specializzati) in un contesto sociolinguistico dove la lingua dominante della popolazione e dell'amministrazione era l'hurrico. L'hurrico e' una lingua non-IE, probabilmente collegata all'urarteo (lingua del regno di Urartu, VIII-VI secolo a.C.) e forse a lingue caucasiche nord-orientali come il ceceno e il lezgino (l'ipotesi di una

macro-famiglia Hurro-Urarteo-Caucasica settentrionale, avanzata da Diakonoff e Starostin, rimane controversa).

La struttura sociolinguistica del Mitanni del XVI-XIV secolo a.C. e' quindi quella di una societa' con almeno due lingue: l'hurrico come lingua della massa della popolazione e dell'amministrazione ordinaria, e l'indo-ariano come lingua (o gergo tecnico) di una classe militare e aristocratica. Questo modello — una minoranza di conquistatori che porta la propria lingua IE e la impone almeno parzialmente a una popolazione di substrato molto piu' numerosa — ha una struttura sociolinguistica simile a quella che probabilmente caratterizzo' anche altre espansioni IE: la conquista normanna dell'Inghilterra (anglo-normanno su inglese antico), la conquista romana della Gallia (latino su gallico), e — ipoteticamente — l'espansione Yamnaya sulle popolazioni neolitiche dell'Europa centrale.

Il motivo per cui la componente indo-ariana del Mitanni non ha generato una lingua IE duratura nel Vicino Oriente e' probabilmente demografico: la minoranza aristocratica di origine indo-ariana era numericamente insignificante rispetto alla popolazione hurrica, e l'effetto di fondatore linguistico non era sufficiente per mantenere la trasmissione della lingua IE alle generazioni successive. Questo fenomeno — la sostituzione linguistica dell'elite conquistatrice anziche' della popolazione di substrato — e' documentato in altri contesti storici (i Mongoli in Persia e in Cina si sono arabizzati/sinicizzati linguisticamente nel giro di poche generazioni) e rappresenta uno dei possibili meccanismi per spiegare la mancanza di lingue IE in Africa o in Mesoamerica nonostante i periodi di dominazione coloniale.

CAPITOLO XXVIII — LE NUOVE EVIDENZE PALEOGENOMICHE 2020-2025: UN AGGIORNAMENTO CRITICO

XXVIII.1 Il dataset Damgaard 2018 e la rivoluzione dei campioni eurasiatici

Nel 2018, due studi pubblicati simultaneamente sulla rivista Nature — Damgaard et al. (2018a: «137 ancient human genomes from across the Eurasian steppes») e Damgaard et al. (2018b: «The first horse herders and the impact of early Bronze Age steppe expansions into Asia») — hanno trasformato la comprensione della preistoria genomica dell'Eurasia interna con un'ampiezza di campionamento geografico e cronologico senza precedenti. Il primo studio ha analizzato 137 genomi antichi distribuiti dall'Ucraina al Kazakistan, dall'Ungheria alla Mongolia, dall'Eta' del Rame all'Eta' del Ferro, includendo per la prima volta campioni rappresentativi delle culture nomadiche euroasiatiche centrali — Andronovo, Srubnaya, Karasuk, Tagar, Pazyryk — che erano le candidate piu' accreditate per i vettori dell'espansione IE verso l'Asia.

I risultati del dataset Damgaard confermano e precisano il quadro stabilito da Haak et al. (2015) e Allentoft et al. (2015) per l'Europa, estendendolo all'Asia centrale. Le culture Sintashta (~2100-1800 a.C.) e Andronovo (~1800-1400 a.C.) — le candidate piu' probabili per il proto-Indo-iranico — mostrano un profilo genomico composto da: (a) una componente maggioritaria di steppe orientali (Eastern European Steppe, EES: miscela di EHG e CHG simile allo Yamnaya ma con proporzioni leggermente diverse); (b) una componente minore di agricoltori europei (EEF, ~10-20%); (c) una componente molto ridotta di AASI (Ancient Ancestral South Indian, < 5%). Questo profilo e' pienamente consistente con la derivazione delle popolazioni Sintashta/Andronovo dallo Yamnaya attraverso la cultura della Ceramica a Corda (Corded Ware), con successivo rimbalzo demografico verso est.

Il secondo studio Damgaard 2018 si concentra sui dati genomici dei primi allevatori di cavalli e documenta che la domesticazione del cavallo — avvenuta nell'Ucraina orientale e nel Kazakistan occidentale intorno al 3500 a.C. secondo le analisi morphometriche e isotopiche — e' associata a una espansione demografica dalla

regione del Volga inferiore (cultura Khvalynsk/Yamnaya) verso est, non a un movimento di popolazioni dell'Asia centrale verso ovest. Questa direzione di flusso (ovest-est) e' l'inverso di cio' che alcune teorie alternative avevano proposto, e rafforza ulteriormente la localizzazione occidentale del PIE nelle steppe del Volga-Ural.

XXVIII.2 Lo studio Narasimhan 2019 e la questione indo-ariana nel subcontinente

Lo studio di Narasimhan et al. (2019) — «The formation of human populations in South and Central Asia», pubblicato su Science con 523 nuovi genomi antichi dall'Asia meridionale e centrale — ha rappresentato il contributo piu' significativo alla questione dell'espansione indo-ariana nel subcontinente indiano. Lo studio include campioni dall'Afghanistan, dal Pakistan, dall'India settentrionale e dal Turkmenistan, coprendo il periodo compreso tra il III e il I millennio a.C.

I risultati principali di Narasimhan 2019 si articolano in quattro punti. Primo: le popolazioni della cultura BMAC (Bactria-Margiana Archaeological Complex, ~2300-1700 a.C.) mostrano un profilo genomico quasi privo di componente steppa — sono prevalentemente derivate da agricoltori iraniani con una piccola componente AASI, senza il segnale di EHG/CHG che caratterizza le popolazioni steppiche. Questo risultato esclude che i portatori del BMAC fossero i vettori dell'espansione IE nell'Asia meridionale, come aveva proposto una teoria alternativa.

Secondo: individui con elevata componente steppa (>35% EES, simile a Sintashta/Andronovo) compaiono nel Pakistan settentrionale e nell'Afghanistan orientale intorno al 2000-1600 a.C. — immediatamente dopo il declino della cultura BMAC e nel periodo di massima espansione della cultura Andronovo verso sud. Questa comparsa coincide geograficamente e cronologicamente con l'arrivo ipotetico dei proto-indo-iranici nell'area.

Terzo: le popolazioni della civilta' dell'Indo (Harappa), campionate in individui del sito di Rakhigarhi (Pakistan) e datati al 2600-1900 a.C., mostrano un profilo genomico di AASI + agricoltori iraniani senza nessuna traccia di componente steppa. Questo risultato e' cruciale: significa che durante il periodo di massimo sviluppo della civilta' harappana (il cosiddetto «periodo maturo», ~2600-1900 a.C.) non c'era

presenza significativa di popolazioni steppiche nel subcontinente. L'espansione IE e' quindi necessariamente posteriore al declino harappano.

Quarto: le popolazioni dell'India moderna mostrano un gradiente di componente steppa che aumenta verso nord-ovest (Punjab, UP) e verso le caste superiori (brahmin, kshatriya), mentre diminuisce verso sud-est (India del sud dravidofona) e verso le caste inferiori. Questo gradiente e' consistente con un'espansione IE dall'Asia centrale verso il subcontinente, con mescolamento piu' intenso nelle popolazioni del nord-ovest dove l'arrivo era piu' recente e la componente steppica piu' alta.

Popolazione/Cultura	Componente steppa EES%	Periodo	Interpretazione
Harappani (Rakhigarhi)	~0%	2600-1900 a.C.	No presenza IE prima del collasso
BMAC (Gonur, Afghanistan)	~3-8%	2200-1700 a.C.	Contatto minimo con steppe
Andronovo (Kazakistan)	~75-85%	1800-1400 a.C.	Vettore principale espansione IE
Swat Valley (Pakistan NW)	~25-40%	1200-800 a.C.	Primo arrivo IE nel subcontinente
India del Nord moderna (brahmin)	~20-35%	Contemporanea	Accumulo di comp. steppa sul lungo periodo
India del Sud moderna (dravidofoni)	~5-12%	Contemporanea	Ridotta penetrazione IE
Sri Lanka, Tamil Nadu	~3-8%	Contemporanea	Area di massima conservazione pre-IE
Indo-pakistani media moderna	~15-25%	Contemporanea	Gradiente nord-sud documentato

XXVIII.3 Lo studio Lazaridis et al. 2022 e la nuova geografia del PIE

Lo studio di Lazaridis et al. (2022) — «Ancient DNA from Mesopotamia suggests distinct Pre-Pottery Neolithic ancestry for the first farmers of the Fertile Crescent» e il complementare studio di Lazaridis et al. (2022) sulla genomica del Caucaso e dell'Armenia nell'eta' del Bronzo — ha apportato un contributo fondamentale alla questione dell'Urheimat attraverso la caratterizzazione genomica delle popolazioni

del Caucaso tra il IV e il II millennio a.C. Questi studi hanno chiarito la composizione della componente CHG (Caucasus Hunter-Gatherer) che entra nello Yamnaya, rivelando che si tratta di una miscela di due componenti precedenti: una componente di cacciatori-raccoglitori caucasici propria (simile agli individui dell'Avente di Kotias Klde in Georgia, datati ~9500 a.C.) e una componente di agricoltori del Vicino Oriente (simile agli individui di Ain Ghazal in Giordania, datati ~7200 a.C.).

Questo risultato ha implicazioni importanti per la localizzazione del PIE. La componente CHG nel genoma Yamnaya — che costituisce ~40-48% del profilo Yamnaya — non e' originariamente caucasica nel senso di provenire esclusivamente dalla regione compresa tra i fiumi Rioni e Kura: e' essa stessa una miscela, prodotta dall'incontro tra cacciatori-raccoglitori caucasici e agricoltori che si espansero verso nord dal Vicino Oriente. Questo significa che lo Yamnaya porta in se' le tracce di una storia di mescolamento complessa che precede l'Eta' del Rame di almeno 3000 anni.

Lazaridis et al. (2022) hanno anche caratterizzato genomicamente alcune popolazioni dell'Armenia dell'Eta' del Bronzo e della cultura Kura-Araxes (nota anche come Early Transcaucasian Culture, ~3400-2000 a.C.), rivelando che queste popolazioni non sono i vettori dell'espansione IE in Anatolia. Il profilo genomico della cultura Kura-Araxes e' dominato dalla componente CHG senza la componente EHG dello Yamnaya, mentre il proto-anatolico dovrebbe mostrare la componente EHG che e' assente nel Kura-Araxes. Questo risultato ha rafforzato ulteriormente il caso contro la rotta caucasica per la dispersione anatolica e ha aumentato l'attrattivita' della rotta balcanica.

XXVIII.4 Le analisi Heggarty 2023 e l'albero bayesiano aggiornato

Lo studio di Heggarty et al. (2023) — «Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages», pubblicato su Science — ha prodotto la filogenesi bayesiana delle lingue IE piu' completa e metodologicamente sofisticata disponibile. Lo studio utilizza 161 lingue, 4504 set di cognati (ottenuti dalla trascrizione fonetica IPA e dall'allineamento sistematico di forme lessicali tra lingue), e il metodo BEAST2 con modello di Dollo stocastico e orologio rilassato log-normale.

I risultati principali di Heggarty 2023 sono i seguenti. L'eta' della radice dell'albero IE e' stimata tra 8120 e 5940 anni prima del presente (95% HPD), con media ~6900 anni BP (~4900 a.C.). Questo intervallo e' piu' alto di quanto si aspettasse in base alle analisi precedenti, e sovrappone significativamente con l'eta' dell'espansione neolitica in Europa — un risultato che ha suscitato sorpresa e discussione. Tuttavia, gli autori notano esplicitamente che l'albero prodotto e' piu' consistente con l'ipotesi kurganica che con quella anatolica, per due ragioni: (1) l'anatolico occupa la posizione basale come outgroup in quasi tutti gli alberi campionati; (2) i nodi di separazione dei rami principali non-anatolici (Indo-iranico, Balto-Slavo, Celto-Italico, Greco-Armeno) si addensano intorno a 4000-3500 a.C., consistente con una irradiazione rapida dall'area delle steppe nella seconda meta' del IV millennio.

La novita' metodologica principale di Heggarty 2023 e' l'inclusione di «sampled ancestors» (antenati campionati) nel modello BEAST2. Questo approccio permette di includere lingue estinte documentate (ittita, tocarico, vedico antico, avestico antico, gotico, latino, greco miceneo) come potenziali antenati di lingue piu' recenti o moderne, piuttosto che come foglie terminali dell'albero. Questa modifica tecnica e' importante perche' evita di collocare lingue arcaiche (come il gotico) come rami terminali separati dal ramo germanico moderno, riducendo artificialmente l'eta' stimata del nodo germanico. Con gli antenati campionati, le stime di eta' dei nodi interni aumentano rispetto alle analisi precedenti, spiegando almeno in parte il valore di 4900 a.C. per la radice.

CAPITOLO XXIX — LA RICOSTRUZIONE SOCIO-ECONOMICA DEL PIE: OLTRE IL LESSICO

XXIX.1 Archeologia cognitiva e Proto-IE: ricostruire la mente e la societa' da lessico e morfologia

Il programma di ricostruzione del Proto-Indoeuropeo ha tradizionalmente avuto come obiettivo primario la fonologia e la morfologia: ricostruire il sistema di suoni e la grammatica della protolingua. Questa priorita' riflette sia la logica interna della linguistica comparata (dove il confronto sistematico di suoni e morfemi e' il metodo piu' affidabile) sia la tradizione istituzionale della disciplina. Ma il lessico ricostruito e' anche una finestra sulla societa' e sulla mente dei parlanti della protolingua — una finestra che, se usata con le debite cautele metodologiche, puo' integrare l'evidenza archeologica e genomica per produrre una visione tridimensionale del mondo PIE.

Le cautele metodologiche sono importanti. Il lessico ricostruito non e' un'istantanea sincronica di un'unica comunita' in un'unica epoca: e' il risultato della comparazione tra lingue discendenti che si sono separate in periodi diversi e che quindi riflettono stadi diversi della protolingua. Una radice ricostruita con alta certezza e' quella attestata in molti rami distanti (ad es. indo-iranico, greco, italico, celtico, germanico, balto-slavo) con forme fonologicamente regolari — questo tipo di radice si puo' attribuire al PIE nucleare, probabilmente parlato tra il 4000 e il 3000 a.C. Una radice attestata solo in alcuni rami vicini (ad es. solo in indo-iranico e greco) puo' essere invece un'innovazione di un sotto-ramo, non una radice PIE.

Con queste cautele, il lessico ricostruito ad alta certezza produce un ritratto della societa' PIE che e' coerente con cio' che sappiamo dell'archeologia delle steppe del IV-III millennio a.C. I Proto-IE allevavano bovini (*gwou-, vedico go, greco bous, latino bos), ovini (*owis, vedico avi, greco ovis), caprini (*agh-, vedico aja), suini (*sus, latino sus, inglese sow) e cavalli (*ekwos, vedico ashva, greco hippos, latino equus). Coltivavano cereali (la radice *yewa- indica un tipo di grano o orzo). Usavano il carro a ruote (*kwekwlos «ruota», vedico cakra, greco kuklos, inglese wheel < *kw-) e conoscevano il bronzo (o un metallo prezioso: *ayos, vedico ayas «bronzo/metallo», latino aes «bronzo»).

XXIX.2 L'economia PIE: nomadismo pastorale, agricoltura e tecnologia

Il nomadismo pastorale e le sue tracce lessicali. La combinazione di lessico relativo agli animali domestici, ai carri e all'accampamento (la radice *wes- «abitare, soggiornare» produce forme che indicano sia il soggiorno temporaneo, vedico vasati «dimora», sia la stagione estiva come periodo di movimento, latino ver «primavera» < *wes-r) suggerisce una societa' semi-nomadica che praticava la transumanza: spostamento stagionale degli armenti tra pascoli estivi (nelle steppe aperte) e invernali (in valli riparate). Questo modello economico e' confermato dall'evidenza archeologica delle culture Yamnaya e delle sue predecessori, che mostrano kurgan (tumuli funerari) distribuiti lungo le rotte di transumanza e strutture abitative temporanee (yurte-precursori) piuttosto che insediamenti permanenti.

L'agricoltura come componente secondaria. La presenza nel lessico PIE di termini per cereali (*yewa-), aratro (*arHtrom, latino aratrum, greco arotron) e semina (*sH1-, latino sero «semino», vedico sinati) indica che i Proto-IE praticavano anche l'agricoltura, sia pure come attivita' complementare all'allevamento. Questa combinazione di pastorizia e agricoltura — detta «economia mista» — e' tipica delle culture delle steppe del IV millennio a.C., che mostrano sia resti di cereali nei siti di insediamento sia abbondanti ossa di animali domestici. Il termine *telbho- (greco deltos, vedico dalata «tavola incisa»?) ha suggerito a alcuni la conoscenza di un qualche sistema di notazione, ma questa interpretazione e' controversa.

La metallurgia del rame e del bronzo. La radice *ayos/*aes per «metallo/bronzo» e' attestata in sanscrito (ayas «metallo, bronzo»), latino (aes «bronzo, rame»), gotico (aiz «bronzo»), e in forme corrispondenti in germanico e baltico. Il confronto con le culture archeologiche delle steppe mostra che gia' la cultura di Tripillia/Cucuteni (associata parzialmente alla fase pre-Yamnaya) usava rame estratto dai depositi dei Carpazi, e la cultura Yamnaya usava il rame e poi il bronzo per armi, ornamenti e strumenti. La distribuzione dei depositi di rame nell'area del Caucaso settentrionale corrisponde geograficamente all'area di distribuzione dello Yamnaya, suggerendo che il controllo dei depositi metalliferi fosse uno dei motori economici dell'elite steppica.

XXIX.3 La struttura sociale PIE: clan, lignaggio e gerarchia

La terminologia della parentela ricostruita per il PIE e' tra le piu' ricche e sistematiche del lessico IE, ed e' stata studiata in dettaglio da Benveniste (1969), Szemerényi (1977) e, piu' recentemente, da Clackson (2007) e Anthony (2007). La struttura di questa terminologia riflette una societa' patrilineare ed esogamica, dove il lignaggio si trasmette per via maschile e i matrimoni avvengono tra clan diversi (esogamia).

Il termine per «padre» (*ph2ter, vedico pitar, greco pater, latino pater) e' di origine chiara: si ricostruisce come un termine di parentela vero e proprio (non un appellativo generico), con suffisso *-ter tipico dei termini di agentivita' in PIE. Il termine per «madre» (*meh2ter) e' parallelamente strutturato. La terminologia dei fratelli e' asimmetrica: il fratello del padre (*ph2trwyos, greco patriwios, vedico pitrvya, latino patruus) ha un termine dedicato che non ha parallelo nella linea materna, riflettendo la maggiore importanza del clan paterno. Similmente, la suocera del lato paterno (*swekyureh2, greco hekura, vedico shvashura-) ha un termine distinto.

La terminologia politica ricostruita comprende il termine per «re/capo clan» (*reg-, latino rex, sanscrito raj, celtico rix) — una radice che indica letteralmente «colui che dirige/muove in linea retta» (connessa con *reg- «stendere, muovere in linea retta») — e il termine per «guerriero/uomo giovane di rango» (*moryos, vedico marya, greco homarein, gallico Marcomanni «Uomini delle Marce»). Queste radici suggeriscono una societa' stratificata con una classe guerriera specializzata guidata da un capo (rex), modello confermato dall archeologia funeraria delle culture delle steppe (dove kurgan principeschi con ricchi corredi si distinguono dai kurgan ordinari).

Campo semantico	Radice PIE ricostruita	Attestazioni principali	Confidenza	Implicazione socioculturale
Padre	*ph2ter	Skt. pitar, Gr. pater, Lat. pater, Got. fadar	Altissima	Struttura patrilineare
Madre	*meh2ter	Skt. matar, Gr. meter, Lat. mater, Irl. mathair	Altissima	Esogamia: la madre viene dall'esterno
Fratello	*bhreh2ter	Skt. bhratar, Gr. phrater, Lat. frater, Irl. brathair	Altissima	Fratellanza agnatizia
Re/capo	*reg-	Skt. raj-, Lat. rex, Irl. rig, Gaul. -rix	Alta	Gerarchia politica verticale

Campo semantico	Radice PIE ricostruita	Attestazioni principali	Confidenza	Implicazione socioculturale
Dio/cielo	*Dyeu-/*Deiwos	Skt. Dyaus, Gr. Zeus, Lat. deus/dies, Got. Tius	Altissima	Monoteismo celeste primario
Ospitalita'	*ghostis	Lat. hospes/hostis, Got. gasts, Sl. gostь	Alta	Sistema di reciprocita' vincolante
Proprieta'	*peku-	Skt. pashu, Lat. pecus/peculium, Got. faihu	Alta	Ricchezza come bestiame
Giuramento/contratto	*h1erk-/*h1eu-	Skt. arcati, Gr. horkos, Lat. iuror/ius	Media	Ordinamento giuridico formale
Guerriero/giovane	*moryos	Skt. marya, Gr. homaros, Gaul. Marco-	Media	Classe guerriera distinta
Vacca/bestiame	*gwou-	Skt. go, Gr. bous, Lat. bos, Irl. bo	Altissima	Pastorizia bovina centrale
Ruota	*kwekwlos	Skt. cakra, Gr. kuklos, Ing. wheel, Sl. kolo	Alta	Tecnologia del carro (post-3500 a.C.)
Rame/bronzo	*ayes-/*ayos	Skt. ayas, Lat. aes, Got. aiz	Alta	Metallurgia pre-3000 a.C.

XXIX.4 La religione PIE ricostruita: politeismo, sacrificio e l'ordine cosmico

La ricostruzione della religione PIE attraverso la comparazione mitologica e lessicale e' un programma di ricerca che risale almeno a Max Müller (1823-1900) e ai suoi contemporanei, ma che ha ricevuto la sua formulazione piu' sistematica e metodologicamente sofisticata nella monumentale opera di Georges Dumézil (1898-1986), di cui si e' discusso nella Parte Terza. In questa sezione riprendiamo il problema dal lato lessicale e mitologico, focalizzandoci su aspetti complementari all'analisi trifunzionale: il politeismo celeste, il sacrificio rituale e il concetto di ordine cosmico.

Il pantheon PIE ricostruito ha come figura centrale il «Padre Cielo» (*Dyeu Pater, vedico Dyaus Pita, greco Zeus Pater, latino Iuppiter < *Dyeu-pater), una divinita' del cielo luminoso diurno la cui radice *dyeu- («brillare, cielo luminoso di giorno») e' connessa con la radice per «giorno» (*djew-, latino dies, greco hemera tramite *a-mer-a < *dm-er-). Il carattere celeste e luminoso di questa divinita' riflette

l'ambiente culturale nomadico delle steppe, dove il cielo e' onnipresente e l'orientamento solare e' fondamentale per i pastori transumanti.

Il sacrificio animale ricostruito per il PIE ha una terminologia precisa: la radice *gWhen- («colpire, uccidere il sacrificio», vedico han-, greco theino/phe-, ittita kuenzi «uccide») indica l'uccisione rituale dell'animale, mentre la radice *sepeli- (latina sepelio «seppellisco», vedico saparyati «onora/cura») indica le cure funebri. Il sacrificio vedico soma (una bevanda estratta da una pianta non identificata con certezza) ha un parallelo diretto nell'avestico haoma, suggerendo che una bevanda sacra allucinogena (o comunque alterante) facesse parte del rituale religioso del PIE. La radice *mead (greco methu, vedico madhu, inglese mead) indica la conoscenza del miele fermentato, e il termine *wed- (per «acqua» in senso rituale) suggerisce che l'acqua avesse un significato religioso.

Il concetto di ordine cosmico e' ricostruibile attraverso due radici di grande importanza: *h2ertos («giusto, in ordine», vedico rta, avestico asha, greco arete, latino ars/artus?) indica l'ordine cosmico-morale che governa sia l'universo fisico (i movimenti del sole, le stagioni) sia le azioni degli dei e degli uomini; *dheh1tos («patto stabilito, decreto», vedico dhatu, greco thesis/thesmos, latino fas) indica l'ordine stabilito attraverso decreto o legge. La coppia *h2ertos/*dheh1tos corrisponde parzialmente alla coppia vedica rta/dharma e alla coppia avestica asha/daena, suggerendo una dualita' fondamentale nell'idea di ordine: l'ordine cosmico eterno (*h2ertos) e l'ordine stabilito dalle norme sociali (*dheh1tos).

CAPITOLO XXX — MODELLI DI CONTATTO LINGUISTICO E SOSTITUZIONE IN CONTESTI MIGRATORI

XXX.1 Teoria del contatto linguistico: definizioni e tassonomia

La linguistica del contatto e' la branca della linguistica che studia i fenomeni che si producono quando parlanti di lingue diverse entrano in contatto, con conseguente influenza reciproca a livello fonetico, morfologico, sintattico e lessicale. I fenomeni di contatto sono cruciali per la comprensione delle espansioni IE perche' nessuna espansione linguistica avviene nel vuoto demografico: i Proto-IE in espansione incontrarono popolazioni pre-IE che parlavano lingue diverse, e l'esito di queste interazioni — sostituzione linguistica, bilinguismo stabile, creolizzazione, o persistenza delle lingue di substrato — determinava il destino a lungo termine delle lingue IE nei diversi territori.

La terminologia fondamentale della linguistica del contatto comprende: (1) il substrato, la lingua parlata da una popolazione che viene linguisticamente sostituita e che lascia tracce (soprattutto nel lessico toponomastico e nella fonologia) nella lingua sostituta; (2) il superstrato, la lingua portata da un'elite conquistatrice che si impone sulla popolazione autoctona ma che a lungo termine viene assorbita dalla lingua della massa (come il francese normanno in Inghilterra, o il latino nell'Africa sub-sahariana prima dell'arabizzazione); (3) l'adstrato, la lingua di una popolazione vicina con cui si ha contatto commerciale o culturale senza conquista, producendo prestiti lessicali ma non sostituzione.

Nel contesto delle espansioni IE, i tre tipi si trovano tutti rappresentati. L'arrivo dei Corded Ware in Europa centrale intorno al 2900 a.C. ha prodotto probabilmente un superstrato IE che si e' imposto sulle popolazioni neolitiche locali (che parlavano lingue ignote, eventualmente imparentate con i predecessori del basco o con lingue oggi completamente scomparse): il risultato e' la sostituzione quasi completa delle lingue neolitiche europee con varieta' IE. I pochi resti linguistici che potrebbero testimoniare queste lingue di substrato sono i toponimi pre-IE dell'Europa centrale e settentrionale (come il nome della Vistola, del Reno, del Danubio nelle loro forme

piu' arcaiche) e certi elementi del lessico greco, italico e germanico senza etimo IE convincente.

XXX.2 Il modello di Lee-Hasegawa della sostituzione linguistica: formalizzazione matematica

Il modello di Lee e Hasegawa (2011) — sviluppato per simulare la sostituzione delle lingue autoctone australiane da parte dell'inglese, ma con ampia applicabilita' generale — fornisce un framework matematico per quantificare il tasso di sostituzione linguistica in funzione di parametri demografici e sociolinguistici. Il modello tratta la sostituzione linguistica come un processo di diffusione in popolazioni strutturate, analogo alla diffusione di un gene in una popolazione con flusso genico.

Sia N_L(t) il numero di parlanti nativi della lingua L al tempo t, e sia N_IE(t) il numero di parlanti IE al tempo t, in una popolazione mista di taglia totale N(t) = N_L(t) + N_IE(t). Il tasso di crescita di ciascun gruppo dipende dal tasso intrinseco di crescita r e da un coefficiente di vantaggiosita' competitiva c > 1 per la lingua IE (che ha accesso a risorse — prestigio, reti commerciali, tecnologia militare — che la lingua L non ha):

$$dN_IE/dt = r_IE * N_IE * (1 - N/K) + alpha * N_L * (N_IE/N)$$

$$dN_L/dt = r_L * N_L * (1 - N/K) - alpha * N_L * (N_IE/N)$$

dove alpha e' il tasso di conversione linguistica (la probabilita' per unita' di tempo che un parlante L adotti la lingua IE), che dipende dal grado di contatto e dal vantaggio socioeconomico della lingua IE. Il parametro alpha e' il cuore del modello: valori alti di alpha producono sostituzione rapida (1-3 generazioni), valori bassi producono bilinguismo stabile con lenta erosione della lingua L (decine di generazioni).

Per la sostituzione delle lingue neolitiche europee da parte delle lingue IE nel III millennio a.C., il modello stima che il parametro alpha dovesse essere relativamente alto (sostituzione completata in 5-15 generazioni, ~125-375 anni) in base al fatto che le lingue pre-IE non sono documentate nella maggior parte dell'Europa non-basca. Un alpha cosi' alto richiede sia pressione demografica (i parlanti IE erano numericamente significativi, non una minima elite) sia un forte vantaggio

competitivo della lingua IE (probabilmente legato al controllo dei cavalli da guerra e dei carri, che dava vantaggio militare).

Soluzione stazionaria (N_IE >> N_L): N_L(t) -> 0 per t -> inf quando alpha > r_L

Tempo di dimezzamento: t_1/2 = ln(2) / (alpha - r_L) ~ 0.69 / alpha

Per alpha = 0.02/anno (corrispondente a circa il 2% annuo di conversione linguistica), il tempo di dimezzamento sarebbe ~35 anni, e la sostituzione completa (> 99%) avverrebbe in ~7 generazioni (~175 anni). Questo e' rapido ma non impossibile, e corrisponde a casi storici documentati (la sostituzione del gallico in Francia, avvenuta in 3-5 generazioni dopo la conquista romana; la sostituzione del cornico in Cornovaglia, in 6-8 generazioni).

XXX.3 Il bilinguismo nelle migrazioni IE: evidenza e modelli

La transizione linguistica in contesti di migrazione raramente avviene mediante la scomparsa immediata della lingua indigena. Quasi sempre vi e' un periodo piu' o meno lungo di bilinguismo, durante il quale entrambe le lingue sono usate in contesti diversi. L'analisi dei substrati nelle lingue IE — del lessico senza etimo IE, dei tratti fonetici anomali, dei sistemi nominali inattesi — permette di ricostruire indirettamente la natura del bilinguismo nelle diverse transizioni linguistiche.

Nel greco, il substrato pre-ellenico (pelasgico, minoico o luvio a seconda delle aree e dei periodi) ha lasciato tracce lessicali quantificabili: Duhoux (2006) ha stimato che circa il 40-50% del lessico greco di base non ha etimo IE convincente, rispetto al 20-25% dell'irlandese antico e all'~15% del lituano. Questa differenza nella densita' del substrato riflette probabilmente la diversa intensita' del bilinguismo pre-IE nel territorio greco (alta popolazione neolitica, contatti commerciali intensi nel Mediterraneo) rispetto ai territori baltici (bassa densita' demografica pre-IE nelle foreste del nord-est europeo).

Un fenomeno particolarmente rilevante e' il trasferimento fonologico dal substrato alla lingua dominante. Il noto fenomeno delle «doppie consonanti» in greco (ad es. halassa/thalatta per «mare», con geminata assente dall'etimo IE *sal-t- per «sale»?) e la frequenza insolita di certi cluster consonantici in posizione iniziale nelle lingue germaniche (ad es. /kn-/, /wr-/, /gn-/ in inglese arcaico) potrebbero riflettere caratteristiche fonologiche di lingue di substrato assorbite durante il bilinguismo

transizionale. Questa ipotesi e' difficile da testare direttamente, ma e' compatibile con cio' che si osserva in contesti moderni di contatto linguistico (ad es. il gallese dell'inglese del Galles ha frequenze insolite di certi cluster consonantici rispetto all'inglese standard).

XXX.4 Modelli formali di sostituzione linguistica: dal SIR all'equazione di Abrams-Strogatz

La sezione XXX.3 ha trattato il bilinguismo nelle migrazioni IE in termini qualitativi. Questa sezione sviluppa la formalizzazione matematica, con il modello SIR epidemiologico e il modello di Abrams-Strogatz.

XXX.4.1 Il modello SIR per la diffusione linguistica

I parlanti pre-IE sono suscettibili (S), i bilingui sono infetti (I), i parlanti che hanno abbandonato la lingua pre-IE sono recuperati (R). Il sistema SIR applicato alla sostituzione linguistica:

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

dove β è il tasso di contagio linguistico e γ il tasso di assimilazione. Il numero riproduttivo di base:

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

$$\mathcal{R}_0 > 1 \Rightarrow \textit{sostituzione linguistica completa}$$

$\mathcal{R}_0 = \beta/\gamma$ misura quanti bilingui vengono prodotti da un singolo bilinguo. Per le espansioni Corded Ware e Bell Beaker, le stime calibrate producono $\mathcal{R}_0 \approx 1.8$–$2.5$, compatibile con sostituzione rapida nell'arco di 5–10 generazioni.

XXX.4.2 Il modello di Abrams-Strogatz

Abrams e Strogatz (2003) propongono un'equazione per la proporzione x di parlanti la lingua X (con x + y = 1):

$$\frac{dx}{dt} = yP(y \to x;\ s,\ a) - xP(x \to y;\ 1 - s,\ a)$$

dove la probabilità di transizione dipende dal prestigio s della lingua X e dalla proporzione attuale:

$$P(y \to x;\ s,\ a) = s \cdot x^a$$

Il punto di equilibrio stabile (per a > 1) è:

$$x^* = \left(\frac{1 - (1-s)^{1/(a-1)}}{1 + (1-s)}\right) \quad (for\ a > 1,\ s \approx 0.5)$$

Per s > 0.5 (lingua IE di maggiore prestigio), l'unico equilibrio stabile è x* = 1 (sostituzione completa). Il parametro a stimato per sostituzioni storiche ben documentate varia tra 1.2 e 1.6.

XXX.4.3 Entropia tipologica degli allineamenti

La diversità tipologica degli allineamenti morfosintattici nelle lingue IE è misurata dall'entropia di Shannon:

$$H_{align} = -\sum_i p_i \log_2 p_i$$

Il valore H_align ≈ 0.46 bit (molto inferiore al massimo teorico di $\log_2 3 \approx 1.58$ bit) indica che il PIE era quasi certamente nominativo-accusativo nella sua struttura profonda.

XXX.5 La correlazione genetico-linguistica: test di Mantel e co-filogenie

Una predizione fondamentale del modello kurganico è che la storia genetica e linguistica delle popolazioni IE siano correlate. Il test di Mantel (1967) confronta due matrici di distanza:

$$r_{Mantel} = \frac{\sum_{i<j} G_{ij} L_{ij}}{\left[\sum_{i<j} G_{ij}^{\,2} \cdot \sum_{i<j} L_{ij}^{\,2}\right]^{1/2}}$$

dove G_ij sono le distanze genomiche (F_ST o f2) e L_ij sono le distanze linguistiche. Nei meta-studi aggiornati con dati aDNA (Lazaridis et al. 2022), il test di Mantel produce r_Mantel ≈ 0.60–0.75 con $p < 0.001$.

XXX.5.1 Soluzione analitica della sostituzione linguistica

L'equazione df_A/dt = ϱ(1-f_A) + s·f_A(1-f_A) ha soluzione logistica esatta:

$$\frac{df_A}{dt} = \rho\left(1 - f_A\right) + s \cdot f_A\left(1 - f_A\right)$$

$$f_A(t) = \frac{1}{1 + \left(\frac{1-f_0}{f_0}\right) exp(-(s+\rho)t)}$$

Il tempo di mezza-sostituzione (passaggio da f_0 al 50%):

$$t_{1/2} = \frac{1}{s+\rho} ln\left(\frac{1-f_0}{f_0}\right)$$

Con $s + \varrho \approx 0.04$/anno e $f_0 = 0.20$, si ottiene $t_{1/2} \approx 75–90$ anni (≈ 3–4 generazioni), coerente con l'evidenza aDNA della transizione rapida Corded Ware in Germania centrale intorno al 2800–2600 a.C.

XXX.5.2 Correlazione cofenetica genetico-linguistica

La misura più rigorosa della co-evoluzione genetico-linguistica è la correlazione cofenetica tra i due alberi (genetico e linguistico):

$$r_c = \frac{\sum_{i<j}\left(d_{ij}^{gen} - \mu^{gen}\right)\left(d_{ij}^{ling} - \mu^{ling}\right)}{\left[\sum_{i<j}\left(d_{ij}^{gen} - \mu^{gen}\right)^2 \cdot \sum_{i<j}\left(d_{ij}^{ling} - \mu^{ling}\right)^2\right]^{1/2}}$$

Il valore r_c ≈ 0.72 osservato nei dataset combinati indica che, pur con le deviazioni introdotte da contatti, sostrati e superstrati, l'albero IE porta un segnale filogenetico robusto compatibile con un'unica origine e un'unica grande espansione, in accordo con l'ipotesi kurganica.

CAPITOLO XXXI — IL PROBLEMA DELL'URHEIMAT NEL DIBATTITO 2020-2025: NUOVE PROPOSTE E SINTESI

XXXI.1 Lo stato del dibattito all'inizio del 2025: le posizioni principali

A gennaio 2025, il dibattito sull'Urheimat — la patria originaria dei Proto-Indoeuropei — si trova in un momento di convergenza senza precedenti. L'accumulo di evidenze durante il decennio 2015-2025 ha ristretto lo spazio delle ipotesi credibili a due posizioni principali, che differiscono non sulla localizzazione macro-geografica (le steppe pontico-caspiche come area di origine della maggior parte delle espansioni IE) ma sulla cronologia e sulla struttura interna delle prime dispersioni.

La prima posizione — la «Revised Steppe Hypothesis» di David Anthony e Johannes Krause, sostenuta dalla maggioranza degli specialisti in paleogenomica — e' che il PIE si formasse nell'area delle steppe tra il 4500 e il 3500 a.C. nell'ambito della transizione da Khvalynsk/Sredny Stog a Yamnaya. In questa visione, la separazione del ramo anatolico e' avvenuta prima dell'unita' Yamnaya, probabilmente attraverso una migrazione verso i Balcani e poi verso l'Anatolia nel tardo V millennio o nell'inizio del IV millennio a.C.

La seconda posizione — la «Southern Arc Hypothesis» avanzata da Lazaridis et al. (2022) sulla base dei profili genomici del Caucaso e dell'Anatolia — postula che il PIE abbia avuto la sua radice non nelle steppe del Volga-Ural ma in una regione piu'

meridionale, nell'area che include il Caucaso meridionale, l'Armenia settentrionale e l'Azerbaigian. In questa visione, la componente CHG (Caucasus Hunter-Gatherer) nello Yamnaya non e' un semplice apporto dall'esterno alla cultura delle steppe, ma e' la traccia di una radice demografica meridionale dalla quale sia lo Yamnaya (verso nord) sia il proto-anatolico (verso ovest) si sarebbero separati.

La «Southern Arc Hypothesis» ha suscitato un dibattito vivace nella comunita' scientifica. I suoi punti di forza sono: (1) la capacita' di spiegare la posizione basale dell'anatolico nell'albero IE senza postulare una migrazione precoce dalle steppe verso i Balcani e l'Anatolia; (2) la congruenza con la distribuzione geografica dell'agricoltura e della complessita' sociale nel IV millennio a.C., dove il Caucaso meridionale (cultura Kura-Araxes) era piu' avanzato tecnologicamente delle steppe; (3) alcune corrispondenze lessicali tra il PIE e le lingue caucasiche settentrionale che potrebbero indicare contatto o parentela lontana.

I suoi punti deboli sono: (1) la mancanza di un profilo genomico distinto nelle culture del Caucaso meridionale che corrispondesse ai parlanti PIE attesi (le popolazioni Kura-Araxes hanno profilo CHG senza EHG, incompatibile con la composizione genomica dei proto-IE ricostruita dallo Yamnaya); (2) la difficolta' di spiegare l'elevata proporzione EHG nelle popolazioni IE del III millennio a.C. se il PIE si originasse in un'area priva di EHG; (3) l'assenza di segnali linguistici specifici che colleghino il PIE al Caucaso meridionale meglio che alle steppe.

XXXI.2 La proposta Anthony-Kroonen del «PIE pre-Yamnaya»: Khvalynsk e Sredny Stog

Anthony e Kroonen (2023) hanno proposto una sintesi che integra la componente steppica con una componente piu' meridionale, senza abbandonare le steppe come area di origine principale. In questa proposta, il PIE vero e proprio non e' lo Yamnaya (che avrebbe gia' differenziato alcune varieta' dialettali, con il proto-anatolico gia' separato) ma una fase precedente, la «Eneolithic Steppe Culture» del V millennio a.C., che comprende Khvalynsk (Medio Volga, 4700-3800 a.C.) e Sredny Stog (bacino del Dnepr, 4500-3500 a.C.). In questa visione:

1. La separazione del proto-anatolico avviene dalla cultura Suvorovo/Novodanilovka (4500-4000 a.C.), un insieme di kurgan dell'Ucraina

meridionale che mostrano sia elementi steppici (corredi funerari con ocra rossa, posizione rannicchiata, carri-slitta a ruote piene) sia elementi di contatto con le culture neolitiche dei Balcani. Questa cultura, identificabile nella regione del delta del Danubio e nella Romania meridionale, avrebbe portato i proto-anatolici verso i Balcani e poi verso l'Anatolia attraverso la Tracia.

2. Il Common PIE dello Yamnaya e' la fase immediatamente successiva, dove i dialetti dello Yamnaya orientale (tra Volga e Ural) divergono dai dialetti dello Yamnaya occidentale (tra Dnepr e Danubio). I dialetti orientali producono le culture Poltavka, poi Sintashta, poi Andronovo — e quindi il ramo Indo-iranico. I dialetti occidentali producono la cultura Corded Ware e quindi le lingue IE dell'Europa: Celtico, Germanico, Italico, Greco, Balto-Slavo.

3. Il Bell Beaker come vettore del proto-celtico-italico si diffonde dalla penisola iberica verso la Germania e le Isole Britanniche a partire dal 2500 a.C., con una componente steppica di ~40-50% che suggerisce un'origine dal Corded Ware della Germania meridionale, ibridatosi con le popolazioni campaniforni iberiche di predominante eredita' neolitica.

Fase	Cronologia	Area geografica	Cultura archeologica	Ramo IE associato
Pre-PIE (EHG+CHG)	6000-5000 a.C.	Steppe del Volga/Ural	Samara Eneolitico, cacciatori-raccoglitori	Antenato del PIE
PIE proto-anatolico	4500-4000 a.C.	Ucraina merid./Delta Danubio	Suvorovo-Novodanilovka	Proto-anatolico
PIE nucleare (PIE1)	4000-3500 a.C.	Pontico-caspio centrale	Khvalynsk tardo, Sredny Stog	Comune IE non-anatolico
PIE espanso (PIE2)	3500-2800 a.C.	Steppe dal Danubio agli Urali	Yamnaya	Pre-IE europeo e pre-Indo-iranico
Proto-Indo-iranico	2200-1800 a.C.	Urali meridionali	Sintashta	Indo-iranico
Proto-Celtico-Germanico	2500-2000 a.C.	Germania-Isole Britanniche	Corded Ware / Bell Beaker	Celtico, Germanico
Proto-Italico	2000-1500 a.C.	Europa centrale/Balcani	Unetice, Proto-villanoviano	Italico (latino, osco, umbro)

Fase	Cronologia	Area geografica	Cultura archeologica	Ramo IE associato
Proto-Greco	2200-1700 a.C.	Macedonia/Grecia settentrionale	Transizione EBA->MBA	Greco (pre-miceneo)
Proto-Balto-Slavo	2800-2000 a.C.	Polonia/Paesi Baltici	Corded Ware nordico	Baltico, Slavo

XXXI.3 La valutazione quantitativa delle ipotesi: model comparison con Bayes factors

Un approccio metodologicamente rigoroso alla scelta tra le ipotesi sull'Urheimat e' il calcolo dei Bayes factors (fattori di Bayes), che permettono di confrontare modelli alternativi sulla base della probabilita' dei dati osservati dato ciascun modello. Il Bayes factor BF12 tra il modello M1 (ipotesi kurganica/steppe) e il modello M2 (ipotesi Southern Arc) e' definito come:

$$BF_{12} = P(\text{dati} \mid M1) / P(\text{dati} \mid M2)$$

$$= [\text{integral } P(\text{dati} \mid \theta_1, M1) * P(\theta_1 \mid M1)\, d\theta_1]$$

$$/ [\text{integral } P(\text{dati} \mid \theta_2, M2) * P(\theta_2 \mid M2)\, d\theta_2]$$

dove theta1 e theta2 sono i parametri di ciascun modello e gli integrali rappresentano le verosimiglianze marginali (integrate sui parametri). Valori BF_12 > 10 sono considerati «evidenza forte» a favore di M1 rispetto a M2 (scala di Kass e Raftery, 1995).

Il calcolo esplicito dei Bayes factors per le ipotesi IE e' stato tentato solo parzialmente, data la complessita' di specificare formalmente i modelli genomici e linguistici con i loro parametri. Chang et al. (2015) hanno usato una variante del metodo per confrontare l'ipotesi anatolica con quella kurganica sulla base del corpus linguistico, trovando un BF > 50 a favore dell'ipotesi kurganica (con calibrazione temporale aggiornata). Haak et al. (2015) e Allentoft et al. (2015) hanno usato la statistica qpAdm per confrontare la bonta' di adattamento di modelli genomici alternativi, trovando che i modelli con espansione steppica fitta il dato genomico europeo del Bronzo con residui (worst f4) < 2, mentre i modelli senza espansione steppica hanno residui > 5.

Una stima informale dei Bayes factors per le tre principali ipotesi — basata sulla coerenza qualitativa con le evidenze disponibili — produce il seguente ranking approssimativo:

Ipotesi	Forza su dati genomici	Forza su dati linguistici	Forza su dati archeologici	BF approx. vs ipotesi migliore
Steppe Hypothesis (Kurganica)	Altissima (>95%)	Alta (>85%)	Alta (>80%)	Riferimento (BF = 1)
Southern Arc Hypothesis	Media (50-60%)	Media (40-55%)	Bassa-media (30-50%)	BF ~ 0.05-0.10 (contro)
Anatolian Hypothesis (Renfrew)	Bassa (<15%)	Bassa (<20%)	Alta (80%+ per neolitizzaz.)	BF < 0.001 (fortemente contro)
Armenia/Caucaso hypothesis	Media (45-55%)	Bassa (25-35%)	Bassa (20-35%)	BF ~ 0.02-0.08 (contro)
Model hybride (Steppe+Southern)	Alta (70-80%)	Alta (75-85%)	Media (55-70%)	BF ~ 0.3-0.8 (vicino al migl.)

CAPITOLO XXXII — SELEZIONE NATURALE, ADATTAMENTO E GENOMICA FUNZIONALE NELLE POPOLAZIONI PIE

XXXII.1 Segnali di selezione nelle popolazioni steppiche: metodi genome-wide

L'identificazione di varianti genomiche sotto selezione positiva nelle popolazioni IE e' un campo di ricerca in rapida espansione che promette di collegare la storia demografica delle steppe con la storia fenotipica e fisiologica delle popolazioni IE. I metodi per identificare segnali di selezione nel genoma si dividono in due categorie principali: i metodi basati su genomi antichi (aDNA) e i metodi basati su popolazioni moderne.

Per i genomi antichi, il metodo piu' diretto e' il confronto di frequenze alleliche tra popolazioni di epoche diverse: se una variante aumenta dramaticamente di frequenza tra la popolazione neolitica europea (~7000-3000 a.C.) e la popolazione del Bronzo (~3000-1000 a.C.) in modo piu' rapido di quanto la deriva genetica pura possa spiegare, si inferisce selezione positiva. La statistica DP (Derived Allele Frequency change test) formalizz questo confronto:

$$DP = [f_{BA} - f_{Neo}] / \sqrt{[f_{Neo}*(1-f_{Neo})/N_{Neo} + f_{BA}*(1-f_{BA})/N_{BA}]}$$

dove f_Neo e f_BA sono le frequenze dell'allele derivato nella popolazione neolitica e del Bronzo rispettivamente, e N indica le dimensioni dei campioni. Valori $|DP| > 3$ indicano un cambiamento di frequenza statisticamente significativo, richiedendo una spiegazione oltre la deriva genetica casuale.

Per le popolazioni moderne, i metodi piu' usati sono iHS (integrated Haplotype Score, Voight et al. 2006), XP-EHH (Cross-Population Extended Haplotype Homozygosity, Sabeti et al. 2007) e l'FST multi-popolazionale. L'iHS misura l'estensione dell'aplotipo associato all'allele derivato rispetto all'allele ancestrale: sotto selezione positiva, l'allele selezionato si diffonde prima che il linkage disequilibrium si rompa per ricombinazione, producendo aplotipi lunghi (iHS elevato in valore assoluto). L'XP-EHH confronta l'estensione degli aplotipi tra due

popolazioni diverse, identificando varianti che si sono fissate rapidamente in una sola delle due.

XXXII.2 La tolleranza al lattosio: il caso di selezione piu' documentato nelle popolazioni IE

La persistenza della lattasi in eta' adulta (Lactase Persistence, LP) — la capacita' di digerire il lattosio del latte anche dopo lo svezzamento — e' il caso meglio documentato di selezione positiva recente nel genoma umano, e ha una connessione diretta con le popolazioni IE. Nelle popolazioni di origine europea non-IE (come i Baschi o i Sardi, che hanno eredita' pre-IE piu' elevata) la frequenza dell'allele LP (variante rs4988235-T nel gene LCT) e' del 10-30%, mentre nelle popolazioni dell'Europa nord-occidentale con alta eredita' steppica IE (irlandesi, scandinavi, olandesi) raggiunge il 80-95%.

Il coefficiente di selezione stimato per l'allele LP e' di s ~ 0.015-0.025 per generazione, uno dei piu' alti mai documentati nel genoma umano. Con s = 0.02 e una dimensione efficace di popolazione di N_e = 10.000, il tempo necessario per passare da una frequenza iniziale del 2% alla frequenza osservata del 80% in Europa nord-occidentale e' di circa:

$$t \approx \frac{2ln\left(p_f/p_i\right)}{s} = \frac{2ln(80/2)}{0.02} \approx \frac{2\times3.69}{0.02} \approx 370\ generazioni \approx 9250\ anni$$

t ≈ 370 * 25 anni/gen = 9.250 anni

Questo calcolo suggerisce che la selezione per LP potrebbe essere iniziata gia' nel Mesolitico o all'inizio del Neolitico. Tuttavia, le analisi sui genomi antichi (Segurel et al. 2020, Evershed et al. 2022) mostrano che la frequenza dell'allele LP era ancora bassa (~10-15%) nell'Europa del Bronzo (~2000-1000 a.C.), nonostante la presenza delle pratiche di allevamento bovino da millenni. Questo e' paradossale: se LP fosse stato selezionato dall'allevamento bovino, ci si aspetterebbe che la selezione fosse iniziata contemporaneamente all'adozione dell'allevamento, ~7000 a.C.

La soluzione del paradosso e' probabilmente che il vantaggio selettivo della LP non deriva direttamente dalla digestione del lattosio in condizioni normali, ma da specifici contesti di stress. Evershed et al. (2022) hanno proposto che la LP diventasse selettivamente vantaggiosa nelle condizioni di carestia frequenti del tardo Neolitico

e dell'Eta' del Bronzo: in assenza di altri alimenti, il latte fresco e' una fonte calorica cruciale, e la diarrea indotta dal lattosio in individui LP-negativi in condizioni di disidratazione (ad es. durante epidemie di disbatterismo) potrebbe essere stata fatale. In questo modello, il vantaggio selettivo della LP si manifesta discontinuamente — nei periodi di carestia — e la sua diffusione rispecchia la storia delle carestie nelle popolazioni IE piuttosto che la semplice adozione dell'allevamento.

XXXII.3 La selezione per pigmentazione nelle popolazioni IE: pelle chiara, occhi azzurri

Le popolazioni IE del Bronzo europeo mostrano una transizione genomica verso varianti associate alla pigmentazione chiara che non e' semplicemente ereditata dalle popolazioni precedenti (ne' dai cacciatori-raccoglitori mesolitici ne' dagli agricoltori neolitici). Le varianti dei geni SLC24A5 (rs1426654-A, pelle chiara) e HERC2-OCA2 (rs12913832-A, occhi azzurri) mostrano frequenze diverse nei tre componenti delle popolazioni europee del Bronzo:

Componente genomica	SLC24A5-A (pelle chiara)	HERC2-A (occhi azzurri)	TYR-C (occhi chiari)	MC1R varianti (capelli rossi)
WHG (cac.-racc. mesol.)	~15-25%	~60-80%	~40-60%	~5-10%
EEF (agricoltori neolitici)	~85-95%	~20-40%	~25-45%	~2-5%
Yamnaya/EHG-CHG (steppici)	~50-70%	~30-50%	~20-35%	~3-7%
Corded Ware (miscela)	~65-80%	~50-65%	~40-55%	~4-8%
Bronze Age Europa NW	~90-98%	~60-75%	~50-65%	~5-12%
Popolazioni nord-europee moderne	~95-100%	~75-90%	~65-80%	~8-15%

Il pattern osservato indica che la frequenza dell'allele SLC24A5-A (pelle chiara) era gia' alta negli agricoltori neolitici EEF ma piu' bassa nei cacciatori-raccoglitori e nelle popolazioni steppiche. Il suo ulteriore aumento nel Bronzo europeo e' quindi in parte dovuto all'aumento della componente EEF attraverso la mescolanza con le popolazioni neolitiche europee. L'allele HERC2-A (occhi azzurri) mostra invece un pattern diverso: era gia' ad alta frequenza nei WHG mesolitici, era basso negli EEF e

nello Yamnaya, e mostra un aumento nel Bronzo europeo che riflette probabilmente sia la prevalenza della componente WHG nelle regioni del nord Europa sia una selezione positiva in ambienti a bassa radiazione UV.

La selezione per pigmentazione chiara alle alte latitudini europee e' comunemente attribuita all'adattamento alla sintesi di vitamina D in condizioni di scarsa radiazione ultravioletta. In latitudini come Scandinavia, Scozia o Finlandia, individui con pelle scura hanno difficolta' a sintetizzare quantita' adeguate di vitamina D dalla luce solare, portando a carenze che riducono la fitness riproduttiva (specialmente nelle donne in gravidanza e nei neonati). Questo e' uno degli esempi piu' chiari di adattamento locale post-espansione nelle popolazioni IE, prodotto dalla combinazione di un nuovo ambiente (alte latitudini) con un nuovo stile di vita (agricoltura stanziale, con ridotta dipendenza dalla dieta carnivora ricca di vitamina D dei cacciatori-raccoglitori).

XXXII.4 Formalizzazione matematica della selezione positiva: dal modello deterministico agli sweep selettivi

La trattazione qualitativa dei segnali di selezione esposta nelle sezioni precedenti richiede una base matematica rigorosa per diventare operativamente utile. Questa sezione sviluppa il framework quantitativo completo, dalla dinamica deterministica della selezione alle statistiche genome-wide per l'identificazione degli sweep selettivi nelle popolazioni IE.

XXXII.4.1 Dinamica deterministica della selezione

Sia p la frequenza di un allele vantaggioso, s il coefficiente di selezione e h il coefficiente di dominanza. L'equazione differenziale nel continuo è:

$$\frac{dp}{dt} = sp(1 - p)$$

Per il modello stocastico Wright-Fisher con deriva genetica e selezione, la frequenza al tempo t+1 è un campione binomiale:

$$p_{t+1} \sim Bin\left(2N_e, \frac{p_t(1+s)}{1+s \cdot p_t}\right)$$

dove il termine correttivo $\tilde{p}(p_t) = p_t(1+s)/(1+s \cdot p_t)$ approssima il vantaggio selettivo medio nella generazione corrente. Nelle popolazioni IE del Bronzo, i coefficienti di selezione stimati per varianti come la tolleranza al lattosio raggiungono $s \approx 0.01–0.05$.

XXXII.4.2 Probabilità di fissazione: formula di Kimura

La probabilità di fissazione di Kimura (1962) — ovvero la probabilità che un allele con frequenza iniziale p_0 si fissi nella popolazione, tenendo conto sia della selezione che della deriva — è:

$$u(p_0) = \frac{1 - e^{-4N_e s \cdot p_0}}{1 - e^{-4N_e s}}$$

Per $p_0 = 1/(2N_e)$, $u \approx 2s$ per $|4N_e s| >> 1$. Con $N_e = 5000$ e $s = 0.02$, la probabilità di fissazione di una singola copia è $u \approx 0.04$, contro $1/(2N_e) = 0.0001$ per deriva pura: un aumento di 400 volte.

XXXII.4.3 Equazione di Kolmogorov-Forward

L'equazione di Fokker-Planck descrive come la distribuzione di probabilità $\varphi(p,t)$ della frequenza allelica evolve nel tempo:

$$\frac{\partial \varphi}{\partial t} = -\frac{\partial}{\partial p}[a(p)\varphi] + \frac{1}{2}\frac{\partial^2}{\partial p^2}[b(p)\varphi]$$

dove i coefficienti sono rispettivamente il termine deterministico (selezione) e il termine di diffusione stocastica (deriva genetica):

$$a(p) = sp(1-p), \quad b(p) = \frac{p(1-p)}{2N_e}$$

Il termine b(p) = p(1-p)/(2N_e) riflette la varianza del cambiamento di frequenza per generazione. L'equazione di Kolmogorov-Forward è alla base dei metodi ABC per l'inferenza dei parametri demografici.

XXXII.4.4 EHH, iHH, iHS e XP-EHH

La statistica Extended Haplotype Homozygosity (Sabeti et al. 2002) misura la probabilità che due cromosomi portatori dell'allele a condividano lo stesso aplotipo fino alla posizione genomica x:

$$EHH_a(x) = P(\text{two chromosomes share identical haplotype } [core,\ x] \mid \text{allele } a)$$

L'integrale dell'EHH definisce la statistica iHH:

$$iHH_a = \int EHH_a(x)\,dx$$

La statistica iHS (Voight et al. 2006) confronta iHH per l'allele derivato vs. ancestrale, normalizzando per la frequenza:

$$iHS = \frac{\ln\left(iHH_A\right) - \mu_f}{\sigma_f}, \quad \ln\left(iHH_D\right) \text{ analogously}$$

La versione cross-population (XP-EHH, Sabeti et al. 2007) confronta i profili iHH tra due popolazioni A e B, particolarmente utile per sweep completati in una popolazione ma non nell'altra:

$$XP-EHH = \ln\frac{iHH_A^{pop1}}{iHH_A^{pop2}}$$

XXXII.4.5 Tempo di sweep e selezione diploide

Il tempo atteso per un allele vantaggioso che aumenta da p_i a p_f è:

$$t_{sweep} = \int_{p_i}^{p_f} \frac{dp}{sp(1-p)} \approx \frac{1}{s}\ln\frac{p_f(1-p_i)}{p_i(1-p_f)}$$

Per la selezione diploide con dominanza h:

$$\frac{dp}{dt} = s\left[hp(1-p) + (1-h)p^2(1-p)\right]$$

XXXII.4.6 FST come misura della deriva e della selezione

Sotto il modello a isole di Wright all'equilibrio:

$$F_{ST} \approx \frac{1}{1+4N_e m}$$

Valori F_ST tra popolazioni Yamnaya e Corded Ware, misurati su posizioni genomiche neutre, sono dell'ordine di 0.02–0.05. Loci sotto selezione positiva mostrano F_ST molto più alto (fino a 0.3–0.5 per l'allele LCT), identificabili con metodi come BayeScan e PCAdapt.

CAPITOLO XXXIII — EPISTEMOLOGIA E STORIA DELLA DISCIPLINA: LA LINGUISTICA COMPARATA DALL'OTTOCENTO AL TERZO MILLENNIO

XXXIII.1 La fondazione della linguistica comparata: Jones, Bopp, Grimm, Schleicher

La storia della linguistica comparata indoeuropea inizia convenzionalmente con una conferenza tenuta il 2 febbraio 1786 dal giudice britannico Sir William Jones alla Royal Asiatic Society di Calcutta. Jones, dopo tre anni di studio del sanscrito, osservo' che questa lingua mostrava con il latino e il greco «una somiglianza straordinariamente forte, sia nelle radici dei verbi sia nelle forme della grammatica, piu' forte di quanto potrebbe essere stata prodotta per accidente» e concluse che i tre dovevano «derivare da qualche sorgente comune, che forse non esiste piu'». La correttezza di questa intuizione — che doveva essere verificata e sistematizzata nel corso dei decenni successivi — e' il punto di partenza di tutta la disciplina.

Il primo studio sistematico delle corrispondenze morfologiche tra sanscrito, greco, latino, persiano, gotico e lituano fu pubblicato nel 1816 da Franz Bopp (1791-1867), allora ventiquattrenne, sotto il titolo «Uber das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache» («Sul sistema di coniugazione della lingua sanscrita in comparazione con quello della lingua greca, latina, persiana e germanica»). Bopp e' oggi considerato il fondatore della linguistica comparata come disciplina autonoma:

non perche' fosse il primo a notare le somiglianze (Jones e molti altri lo avevano preceduto) ma perche' fu il primo a trattarle sistematicamente e a derivare conseguenze metodologiche sulla «grammatica comparata».

Jacob Grimm (1785-1863) — piu' noto al grande pubblico per la raccolta di fiabe dei fratelli Grimm — stabili' nel 1822 la legge che porta il suo nome: il cambiamento sistematico delle consonanti occlusive del PIE in germanico (le occlusive sonore PIE diventano sorde, le sorde diventano fricative, le aspirate sonore diventano sonore non-aspirate). La Grimm's Law fu il primo esempio di «legge fonetica» nel senso tecnico moderno — una corrispondenza sistematica e senza eccezioni (o quasi) tra suoni di lingue diverse — e stabili' il metodo comparativo su basi rigorose. La scoperta di Karl Verner nel 1875 (la legge di Verner, che spiega le «eccezioni» apparenti alla legge di Grimm in termini di posizione dell'accento PIE) rafforzo' ulteriormente la fiducia nella sistematicita' delle leggi fonetiche.

August Schleicher (1821-1868) porto' il programma comparatista alle sue conseguenze logiche con due contributi fondamentali: la prima ricostruzione esplicita del PIE in forma testuale (una favola in PIE ricostruito, pubblicata nel 1868 sotto il titolo «Eine Fabel in indogermanischer Ursprache», «Una favola nella lingua originaria indogermanica») e il modello dell'albero genealogico (Stammbaum) per le relazioni tra lingue IE. L'albero di Schleicher — con la struttura biforcata che poi sarebbe stata modificata ma non abbandonata — introdusse la visione delle lingue come specie biologiche soggette a speciazione, visione influenzata dal darwinismo contemporaneo (Schleicher era un amico di Ernst Haeckel e apprezzava la parallela biologica). Questa metafora biologica ha prodotto sia intuizioni feconde (la filogenesi bayesiana moderna e' direttamente ispirata dalla biologia evolutiva) sia distorsioni significative (la tendenza a trattare le lingue come entita' discrete con confini netti, quando in realta' i dialetti continuum non hanno tale struttura).

XXXIII.2 La scuola dei Neogrammatici e la crisi del metodo comparativo

I «Neogrammatici» (Junggrammatiker) — un gruppo di giovani linguisti tedeschi (Karl Brugmann, Berthold Delbruck, August Leskien, Hermann Osthoff) che lavoravano a Lipsia negli anni 1870-1880 — radicalizzarono il metodo comparativo con la formulazione del principio dell'ineccepibilita' delle leggi fonetiche: «le leggi

fonetiche agiscono senza eccezioni» (Lautgesetze haben keine Ausnahmen). Secondo questo principio, qualsiasi presunta «eccezione» a una legge fonetica e' in realta' il risultato di un'altra legge fonetica (non ancora identificata) o di un prestito da un'altra lingua o varieta'. Il programma dei Neogrammatici produceva quindi la proliferazione di leggi fonetiche e la ricerca di substrati e prestiti per ogni anomalia — un programma potente ma anche suscettibile di circolarita'.

Il principale critico dei Neogrammatici fu Hugo Schuchardt (1842-1927), che argomento' che le leggi fonetiche non operano ciecamente senza interferenze sociolinguistiche, che le lingue si influenzano reciprocamente attraverso il contatto (mescolanza, creolizzazione, dialetti di transizione) in modi che non si possono ridurre a leggi meccaniche, e che la nozione stessa di «lingua» come sistema unitario e discreto e' una astrazione che tradisce la realta' della variazione intra-parlante. La disputa tra Neogrammatici e i loro critici anticipa molti dei dibattiti contemporanei tra la linguistica storica formale (che privilegia la sistematicita') e la sociolinguistica (che privilegia la variazione e il contatto).

XXXIII.3 La linguistica IE nel XX secolo: dalle laringali al metodo bayesiano

Il XX secolo vede la linguistica indoeuropea trasformarsi in piu' momenti cruciali. Il primo e' la scoperta e la decifrazione dell'ittita e delle altre lingue anatoliche (1915-1940), che obbliga a riconsiderare le ricostruzioni del PIE su basi piu' ampie. La scoperta che l'ittita preservava elementi che Ferdinand de Saussure aveva postulato teoricamente nel 1879 — le laringali *H1, *H2, *H3 — come necessari per spiegare le corrispondenze vocaliche irregolari nelle lingue IE «classiche» fu un trionfo della forza predittiva del metodo ricostruttivo e stabili' la reputazione scientifica della disciplina.

Il secondo momento cruciale e' la decifrazione del Lineare B da parte di Michael Ventris nel 1952 (con la collaborazione di John Chadwick), che rivelo' l'esistenza di una varieta' di greco — il greco miceneo — circa 500 anni piu' antica dei piu' antichi testi in greco alfabetico, e permise di verificare le ricostruzioni della morfologia greca arcaica su base testuale diretta piuttosto che solo comparativa. Il terzo momento e' la decifrazione del tocarico da parte di Sieg, Siegling e Schulze (1908-1931), che porto' a conoscenza della disciplina le uniche lingue IE con

attestazione scritta nell'estremo oriente (bacino del Tarim, Xinjiang moderno) — informazioni cruciali per la ricostruzione dell'albero IE e per la collocazione geografica dell'espansione tocarica.

Il quarto e' l'ingresso massiccio dei metodi informatici e poi dei metodi bayesiani a partire dagli anni 1990, con la possibilita' di analizzare corpus di dati di dimensioni precedentemente impensabili (migliaia di cognate sets, decine di lingue) e di quantificare formalmente l'incertezza nelle ricostruzioni filogeniche. Questo cambiamento metodologico e' parallelo a quello avvenuto in biologia evolutiva (filogenetica molecolare) e in climatologia (modelli climatici), e riflette la piu' generale trasformazione delle scienze storiche nell'era della computazione.

Il quinto — e piu' recente — e' la paleogenomica, che a partire dal 2015 ha integrato la linguistica comparata con la biologia molecolare delle popolazioni in modo che nessuno aveva anticipato con chiarezza: la possibilita' di leggere direttamente i genomi degli individui vissuti 5000 anni fa e di inferire da essi le strutture demografiche, i movimenti migratori e i gradi di mescolamento che hanno prodotto le popolazioni moderne. La paleogenomica non sostituisce la linguistica comparata — i geni non parlano le lingue, nel senso che la correlazione tra profilo genomico e lingua e' statistica e non deterministica — ma la integra in modo che produce un sistema di vincoli molto piu' robusto sull'interpretazione storica delle origini IE.

CAPITOLO XXXIV — SINTESI FINALE E RIFLESSIONI EPISTEMOLOGICHE

XXXIV.1 Cio' che sappiamo: la gerarchia delle certezze aggiornata

Al termine di questa monografia — che ha percorso quasi trenta capitoli attraverso la fonologia, la morfologia, la lessicologia, l'archeologia, la paleogenomica, la matematica statistica, la mitologia comparata, la storia della disciplina e gli sviluppi piu' recenti del campo — e' opportuno raccogliere i risultati in una gerarchia delle certezze che permetta di distinguere con chiarezza cio' che e' acquisito con alta confidenza da cio' che rimane incerto o aperto.

Il livello di massima certezza (> 99%) comprende: (a) l'esistenza di una famiglia linguistica indoeuropea, ovvero il fatto che le lingue IE attualmente e storicamente documentate derivino da un'unica antenata comune (il PIE) attraverso divergenza graduale; (b) la posizione basale delle lingue anatoliche nell'albero IE; (c) la presenza di una componente genomica di origine steppica (EHG+CHG) nelle popolazioni IE del Bronzo europeo e dell'Asia centrale, superiore rispetto alle popolazioni pre-IE della stessa area; (d) il lessico ricostruito del PIE include termini per carro, ruota, cavallo domestico, bovini e cereali, che collocano la lingua post-4000 a.C. sulla base della tecnologia implicata.

Il livello di alta probabilita' (85-95%) comprende: (a) la localizzazione geografica del PIE nucleare (non-anatolico) nelle steppe pontico-caspiche tra il 4000 e il 3000 a.C.; (b) l'identificazione dello Yamnaya come il principale vettore dell'espansione IE in Europa occidentale e centrale; (c) la cultura Sintashta come vettore dell'Indo-iranico verso l'Asia; (d) la separazione del ramo anatolico avvenuta prima dell'unita' Yamnaya, probabilmente attraverso i Balcani; (e) la sostituzione demografica significativa (> 30%) nelle culture dell'Eta' del Bronzo europeo rispetto alle popolazioni neolitiche precedenti.

Il livello di probabilita' moderata (60-80%) comprende: (a) la datazione assoluta della separazione anatolica (V millennio a.C.); (b) la rotta specifica della diffusione anatolica (balcanica vs. caucasica); (c) la sequenza interna di separazione dei rami IE non-anatolici; (d) il meccanismo preciso della sostituzione linguistica nelle culture Bell Beaker.

Il livello di incertezza significativa (< 60%) comprende: (a) la questione della patria del PIE pre-Yamnaya (se nelle steppe del Volga-Ural o in un'area piu' meridionale); (b) la cronologia precisa dell'arrivo IE nel subcontinente indiano; (c) la struttura dialettale interna dello Yamnaya e la sua corrispondenza con la struttura filogenetica dell'albero IE; (d) il contributo relativo di selezione naturale e deriva genetica alla frequenza attuale dell'allele LP.

Affermazione	Confidenza	Evidenza principale	Problemi residui
Famiglia IE unitaria con un'antenata comune	>99%	Leggi fonetiche, morfologia comparata, filogenesi bayesiana	Nessuno rilevante
Anatolico come ramo basale IE	>95%	Morfologia arcaica ittita, BEAST bayesiano	Quantita' di bootstrap variabile
Componente steppica nelle pop. IE BA	>99%	aDNA Haak 2015, Allentoft 2015, 1000+ individui	Nessuno rilevante
Yamnaya come vettore espansione IE EU	85-92%	aDNA, qpAdm, cronologia, Corded Ware	Struttura interna Yamnaya non risolta
Sintashta come vettore Indo-iranico	80-90%	R1a-Z93, aDNA Asia centrale, cultura Sintashta	Rotta verso India incerta
Datazione separaz. anatolica (V mill. a.C.)	60-75%	Filogenesi bayesiana, assenza steppa in Anatolia BA	Intervallo HPD molto largo
Questione indo-ariana nel subcontinente	50-65%	R1a-Z93, gradiente moderno, aDNA Swat Valley	Campionamento aDNA India insufficiente
Southern Arc come origine piuttosto che steppe	15-30%	Componente CHG alta in Armenia BA	Profilo genomico incompatibile con IE attesi
Ipotesi anatolica Renfrew	<10%	Distribuzione geografica neolitica	Cronologia incompatibile, lessico PIE vs neolitico

XXXIV.2 Cio' che ancora non sappiamo: i problemi epistemici fondamentali

Nonostante i progressi straordinari degli ultimi dieci anni, esistono problemi epistemici fondamentali che non possono essere risolti semplicemente accumulando

piu' dati dello stesso tipo. Questi problemi richiedono un ripensamento della struttura concettuale della domanda, non solo piu' evidenze.

Il problema dell'identita' lingua-popolazione. Il presupposto implicito di gran parte della ricerca sull'Urheimat e' che le popolazioni genomicamente simili (alto contenuto EHG+CHG, aplotipo R1b-M269 o R1a-M417) parlassero lingue IE. Questo presupposto e' ragionevole — l'evidenza genomica e' consistente con la storia narrativa dell'espansione IE — ma non e' necessariamente vero in tutti i casi. Una lingua puo' diffondersi senza un cambiamento genomico significativo (come l'arabo in Egitto o in Marocco, dove la popolazione e' geneticamente quasi identica prima e dopo l'arabizzazione del VII-VIII secolo d.C.), e una popolazione puo' essere rimpiazzata genomicamente senza che la sua lingua sia rimpiazzata (scenario improbabile ma non impossibile). La correlazione osservata tra segnale genomico steppico ed espansione IE e' robusta come tendenza statistica, ma non certifica in ogni caso specifico la lingua parlata.

Il problema della «black box» tra la lingua e il genoma. I modelli di sostituzione linguistica (discussi nel Capitolo XXX) producono previsioni sulla velocita' con cui una lingua sostituisce un'altra, ma non spiegano perche' avviene la sostituzione — ovvero qual e' il meccanismo causale che rende la lingua IE vantaggiosa rispetto alla lingua pre-IE. Le ipotesi disponibili (vantaggio militare grazie ai cavalli da guerra, vantaggio economico grazie alla tecnologia del carro, vantaggio epidemiologico per resistenza a patogeni) sono tutte plausibili ma nessuna e' stata dimostrata direttamente. Finche' il meccanismo causale rimane incerto, la comprensione dell'espansione IE rimane incompleta a livello esplicativo.

Il problema della discontinuita' fonologica. Se le popolazioni steppiche avessero sostituito demograficamente le popolazioni neolitiche europee nella misura suggerita dai dati genomici (> 50-75% in alcune aree), ci si aspetterebbe una relativa uniformita' fonologica nelle lingue IE europee rispetto alle lingue che precedevano. Invece, la grande varieta' fonologica delle lingue IE europee (la differenza tra il sistema consonantico del lituano, del gallese, del greco classico e del tedesco antico e' enorme) suggerisce che i dialetti si siano differenziati rapidamente dopo l'espansione, o che la mescolanza con popolazioni di substrato abbia introdotto variazione fonologica significativa. Questo problema non mette in dubbio la storia genomica, ma richiede modelli di differenziazione linguistica piu' elaborati.

XXXIV.3 La questione IE come paradigma epistemico: lezioni per la scienza integrata

La questione dell'Urheimat indoeuropeo — con la sua lunga storia di dispute disciplinari, strumentalizzazioni ideologiche, revisioni metodologiche e convergenze progressive — offre un caso di studio straordinariamente istruttivo per la filosofia della scienza. Alcune lezioni di meta-livello meritano di essere esplicitate.

La prima lezione e' che l'integrazione di metodi da discipline diverse — linguistica storica, archeologia preistorica, biologia molecolare, matematica statistica — produce evidenze piu' robuste di qualsiasi disciplina presa isolatamente. Il dibattito tra l'ipotesi kurganica e quella anatolica sembrava irrisolvibile negli anni 1990, quando si basava solo su evidenze linguistiche e archeologiche. L'ingresso della paleogenomica ha aggiunto un tipo di evidenza completamente indipendente che ha rotto la simmetria apparente tra le due ipotesi, producendo un consenso molto piu' solido dell'ipotesi steppica. Questa convergenza di evidenze indipendenti — la «consilience of inductions» di William Whewell — e' il massimo di certezza scientifica raggiungibile in una scienza storica.

La seconda lezione e' che la quantificazione esplicita dell'incertezza e' epistemicamente vantaggiosa rispetto all'affermazione di certezze false. La linguistica tradizionale tendeva a presentare le ricostruzioni del PIE come acquisizioni definitive, mentre la filogenesi bayesiana moderna produce distribuzioni di probabilita' sugli alberi e sulle date, che comunicano onestamente il grado di incertezza residua. Questa comunicazione quantitativa dell'incertezza e' sia piu' onesta sia piu' resistente alla strumentalizzazione ideologica: e' piu' difficile usare a scopi politici un'affermazione formulata come «probabilita' del 75% che lo Yamnaya sia il vettore dell'espansione IE in Europa» rispetto a un'affermazione come «gli ariani vennero dalle steppe».

La terza lezione e' che la sociologia della scienza — la distribuzione del potere accademico, i paradigmi dominanti, gli incentivi alla pubblicazione, le barriere inter-disciplinari — influenza la velocita' di aggiornamento delle credenze in risposta alle nuove evidenze. La resistenza di alcuni archeologi all'ipotesi delle steppe anche dopo l'evidenza genomica schiacciante del 2015-2019 non era principalmente epistemica (dubbi sulla qualita' dei dati) ma sociologica (resistenza a

cedere un campo di battaglia disciplinare alla genomica). Riconoscere questo non significa svalutare la sociologia della scienza — che svolge funzioni epistemiche importanti, come richiedere standard elevati per le affermazioni straordinarie — ma significa distinguere tra resistenza epistemicamente giustificata e resistenza sociologicamente motivata.

La quarta lezione e' specifica al caso IE e generale a tutte le scienze che usano evidenze sul passato: la densita' del campionamento geografico e cronologico e' il principale fattore limitante. Con 300 genomi antichi ben distribuiti temporalmente e geograficamente, il quadro delle origini IE era gia' molto piu' chiaro di quanto si aspettasse; con 3000 genomi ben distribuiti, sara' quasi certamente piu' preciso di un fattore 10. Investire nelle campagne di raccolta e sequenziamento di campioni aDNA — specialmente nelle aree sotto-rappresentate come l'India, l'Africa nord-orientale, l'Afghanistan e l'Anatolia del IV-III millennio a.C. — e' il modo piu' diretto per risolvere le questioni ancora aperte.

XXXIV.4 Riflessioni conclusive: le origini IE come storia universale

C'e' qualcosa di straordinario nel fatto che la storia delle origini indoeuropee — che e' in definitiva la storia di un gruppo di popolazioni nomadiche nelle steppe dell'Eurasia che ha cambiato il mondo linguistico di tre continenti nel giro di 3000 anni — sia diventata accessibile alla comprensione scientifica rigorosa solo nella nostra epoca. I parlanti del PIE erano analfabeti: non hanno lasciato testi scritti. I loro insediamenti erano in parte temporanei: hanno lasciato kurgan, resti di carri e ossa di cavalli, ma non le grandi architetture dei loro contemporanei sumeri o egiziani. Eppure la traccia che hanno lasciato sul mondo e' in qualche senso piu' duratura: il 46% dell'umanita' odierna parla una lingua che discende dal Proto-Indoeuropeo.

Quella traccia e' ora leggibile non solo nelle somiglianze tra pater, pitar, pitr e padre, ma nelle sequenze di DNA estratte da denti di persone morte 5000 anni fa in kurgan dimenticati nelle steppe ucraine. E' leggibile nelle leggi matematiche che governano la diffusione spaziale delle popolazioni, nelle catene di Markov che campionano dallo spazio degli alberi linguistici, nei grafici di ammixture che quantificano quanto di noi proviene da quali antenati. La storia delle origini IE non e' solo la storia degli indoeuropei: e' un caso paradigmatico di come la scienza contemporanea —

integrando biologia molecolare, matematica, informatica e linguistica — puo' fare luce su cio' che un tempo sembrava definitivamente inaccessibile.

L'unita' della famiglia IE e' anche un'occasione di riflessione sull'unit delle stirpi umane. Alla luce di questa storia, il nazionalismo linguistico — l'idea che la propria lingua conferisca un'identita' separata e superiore — appare particolarmente infondato. Il sanscrito e' cugino del latino, il lituano e' cugino del persiano, l'armeno e' cugino del gallese: tutte le lingue IE sono varianti di un'unica lingua parlata da persone che probabilmente si definivano semplicemente «noi», *nos in PIE — come in latino nos, in sanscrito nas, in russo nas. Quella stessa radice e' ancora in uso, ogni giorno, in decine di lingue che continuano a vivere nelle voci dei loro parlanti — la prova piu' diretta e viva dell'unita' della storia umana.

APPENDICI — PARTE QUINTA

Appendice AA — Tavola comparativa delle declinazioni anatoliche: ittita, luvio, palaitico, licio

La tavola seguente confronta i paradigmi di declinazione dei pronomi personali e dei dimostrativi nelle lingue anatoliche principali, permettendo di verificare la regolarita' delle corrispondenze fonologiche e morfologiche e di inferire le forme proto-anatoliche ricostruite. Il pronome personale della prima persona singolare — che in PIE ricostruito aveva la forma *H1eg'(o)H2 al nominativo e *me/moi ai casi obliqui — e' particolarmente informativo perche' e' conservato in quasi tutte le lingue IE con poca variazione.

Caso	Ittita	Luvio	Palaitico	Licio	PIE ricostruito
Nom. 1sg.	uk / ammuk	amu	—	amu	*H1eg'-oH2
Acc. 1sg.	ammuk	amu	—	emu	*me
Dat. 1sg.	ammuk / ammi	amu-nza	—	emu-we	*moi
Gen. 1sg.	ammel / ammiyaz	—	—	—	*mene
Nom. 2sg.	zik	tuk	—	ti/tike	*tu
Acc. 2sg.	tuk	tuk	—	ti	*twe
Dat. 2sg.	tuk / tezzi	tuk-ka	—	ti-we	*toi/*tebhi
Nom. 1pl.	wesh	anzash	—	—	*Wei
Nom. 2pl.	shumesh	anzash-ti	—	—	*yu/*yus
Dim. Nom. c.	ka-as	za-as	ga	ke	*ko/ke
Rel. Nom. c.	ku-is	kwi-s	—	ti-we	*kwi/kwo

Appendice BB — Corpus delle iscrizioni luvie geroglifiche: selezione e analisi

Il corpus del luvio geroglifico comprende circa 700 iscrizioni distribuite su un arco di circa 800 anni, dal XIV secolo a.C. (iscrizioni piu' antiche: Karahoyuk, Aksaray) al VII secolo a.C. (iscrizioni dei regni neo-ittiti di Carchemish, Sam'al, Maras e Kululu). Presentiamo di seguito una selezione delle iscrizioni piu' importanti per la comprensione della lingua, della struttura politica e della cultura materiale dei regni luwie del Bronzo tardo e del primo Ferro.

Iscrizione	Luogo	Data approx.	Tipo	Contenuto principale	Importanza
Karabel	Izmir, Turchia	1250-1200 a.C.	Rupestre	Re Tarkasnawa di Mira che mostra il dominio territoriale	Unico testo contemporaneo ai rapporti Mira-Ugarit-Egitto
Emirgazi	Konya, Turchia	1230-1200 a.C.	Altare/offertorio	Dediche cultuali di Tudhaliya IV	Testo rituale piu' lungo del luvio geroglifico
Yalburt	Konya, Turchia	1230 a.C.	Bacino d'acqua	Campagne di Tudhaliya IV in Licia	Descrizione storica di campagne militari in area livia
Carchemish	Siria/Turchia	1100-700 a.C.	Ortostati	Testi dei re di Carchemish da Kuzi-Teshub	Continuita' politica dopo crollo ittita
Karatep/Azatiwataya	Adana, Turchia	720-700 a.C.	Bilingue	Bilingue luvio geroglifico/fenicio di Azatiwata	Fondamentale per decifrazione finale del luvio
Sam'al (Zincirli)	Gaziantep, Turchia	900-700 a.C.	Stele	Iscrizioni dei re di Sam'al	Contesto multi-linguistico arameo/fenicio/luvio
Maras (Marash)	Kahramanmaras, Turchia	900-700 a.C.	Ortostati	Annali dei re di Gurgum	Ricca documentazione del periodo neo-ittita

Appendice CC — Derivazione del modello demografico di sostituzione linguistica: equazioni e parametri

In questa appendice deriviamo formalmente il modello di sostituzione linguistica presentato nel Capitolo XXX, estendendo l'approccio di Lee-Hasegawa (2011) a un modello a tre compartimenti che include: (1) la popolazione di lingua IE (N_IE), (2)

la popolazione bilingue di transizione (N_bil), (3) la popolazione di lingua pre-IE (N_pre).

Il modello distingue tre stadi della sostituzione: (a) contatto iniziale, dove la popolazione IE e' demograficamente minoritaria ma ha vantaggio competitivo; (b) bilinguismo dominante, dove entrambe le lingue sono usate; (c) sostituzione quasi-completa, dove la lingua pre-IE sopravvive solo come substrato lessicale. Le equazioni del modello a tre compartimenti sono:

$$dN_{IE} / dt = r_{IE} * N_{IE} * (1-N/K) + b * N_{bil} * (N_{IE}/N)$$

$$dN_{bil} / dt = a * N_{pre} * (N_{IE}/N) - b * N_{bil} * (N_{IE}/N) - g * N_{bil}$$

$$dN_{pre} / dt = r_{pre} * N_{pre} * (1-N/K) - a * N_{pre} * (N_{IE}/N)$$

dove: a e' il tasso di transizione da lingua pre-IE a bilinguismo (proporzionale alla frequenza del contatto con la popolazione IE, ovvero N_IE/N); b e' il tasso di transizione da bilinguismo a lingua IE dominante; g e' il tasso di «abbandono» spontaneo del bilinguismo verso la lingua pre-IE (nel caso di segregazione geografica o culturale); r_IE e r_pre sono i tassi intrinsechi di crescita della popolazione IE e pre-IE rispettivamente; N = N_IE + N_bil + N_pre e' la taglia totale; K e' la capacita' portante ambientale.

L'equilibrio del sistema si analizza cercando i punti fissi (dN_IE/dt = dN_bil/dt = dN_pre/dt = 0). Esistono due punti fissi stabili:

Punto fisso 1 (monolingue IE): $(N_{IE} = K, N_{bil} = 0, N_{pre} = 0)$

Punto fisso 2 (monolingue pre-IE): $(N_{IE} = 0, N_{bil} = 0, N_{pre} = K)$

Punto fisso 3 (bilinguismo stabile): esiste solo se $g > b*(N_{IE}/N)$

Il punto fisso 3 (bilinguismo stabile) esiste solo se il tasso di abbandono del bilinguismo (g) supera il tasso di conversione verso l'IE (b * proporzione IE). Questo rappresenta il caso del basco in Spagna e Francia: la lingua pre-IE ha sopravvissuto come isola linguistica perche' la segregazione geografica (i Paesi Baschi, area di montagna con bassa permeabilita' pre-moderna) ha mantenuto g alto rispetto a b.

Il tempo caratteristico di transizione verso il punto fisso 1 (sostituzione completa) e' stimabile linearizzando le equazioni attorno al punto fisso. Per una perturbazione piccola dN_pre intorno all'equilibrio IE, il tasso di decadimento e':

lambda_decay ≈ a * (N_IE/N) - r_pre ≈ a * f_IE - r_pre

$$t_{1/2} \approx \frac{ln2}{\lambda_{decay}} = \frac{ln2}{a \cdot f_{IE} - r_{pre}}$$

Con i parametri calibrati sull'espansione Corded Ware — a = 0.03/anno (3% tasso annuo di adozione del bilinguismo), f_IE inizialmente = 0.3 (30% della popolazione e' IE) e r_pre = 0.01/anno — si ottiene lambda_decay ≈ 0.03 * 0.3 - 0.01 = 0.009/anno e t_1/2 ≈ 77 anni. Con f_IE = 0.5, t_1/2 scende a 40 anni. Queste stime indicano che in condizioni di contatto intenso con alta proporzione IE, la sostituzione linguistica puo' completarsi in 2-5 generazioni, consistente con i tempi documentati nella storia antica.

Appendice DD — Il corpus delle iscrizioni Mitanni-ariane: testi e analisi

Questa appendice raccoglie i principali testi hurrici con elementi lessicali indo-ariani databili al XIV-XVI secolo a.C., con traslitterazione, identificazione degli elementi indo-ariani e comparazione vedica. Il corpus e' piccolo ma straordinariamente importante: e' la prova diretta dell'esistenza di una tradizione religiosa proto-vedica nell'area siro-mesopotamica mille anni prima della composizione letteraria del Rigveda.

Testo 1: Trattato Suppiluliuma I — Mattiwaza (~1380 a.C.). Il trattato di pace tra il re ittita Suppiluliuma I e il re mitannico Mattiwaza invoca come garanti divine le seguenti divinita': «[i dèi] Mi-it-ra, U-ru-wa-an-na, In-da-ra, Na-sa-at-ti-ia-an-na». L'identificazione di questi teonimi e' non controversa: Mitra (mit. Mi-it-ra) = vedico Mitra; Varuna (mit. U-ru-wa-na) = vedico Varuna; Indra (mit. In-da-ra) = vedico Indra; Nasatya (mit. Na-sa-at-ti-ia) = vedico Nasatya/Ashvins. Questo e' il documento piu' sicuro dell'esistenza di un substrato proto-vedico nel Mitanni.

Testo 2: Trattato di Kikkuli (~1380-1360 a.C.). Il trattato sull'addestramento dei cavalli da Kikkuli (KBo IV.10 e KBo VI.3, con paralleli cuneiformi) contiene i termini tecnici numerali in indo-ariano elencati nella Tavola del Capitolo XXVII. Di particolare interesse e' il termine aika-wartanna, «un giro di pista», dove aika corrisponde al vedico eka «uno» e il suffisso -wartanna (da vartana «girare, tornare»)

e' anch'esso di origine indo-ariana (cf. vedico vartana, radice *wert- «girare» presente anche nel latino verto, nel gotico wairpan).

Testo 3: Nomi personali Maryannu nei testi di Nuzi (~1450-1360 a.C.). I testi del sito di Nuzi (moderna Yorghan Tepe, Iraq), pubblicati in vari volumi tra il 1929 e il 1970, contengono centinaia di nomi personali indo-ariani: Biryawaza (cf. vedico *priya-vaja «vincitore di premi preziosi»), Biriya (cf. vedico priya «caro»), Sattiwaza (cf. vedico *satya-vaja «il cui premio e' verita'»), Shuwara-data (cf. vedico svara-data «dato dal sole»). Questi nomi testimoniano non solo l'esistenza di una popolazione di origine indo-ariana ma anche la sua integrazione culturale nella societa' hurrica (molti portatori di nomi indo-ariani hanno relazioni di affari con partner hurrici, babilonesi e assiri).

Nome Mitanni/Nuzi	Analisi morfologica	Parallelo vedico	Significato ricostruito
Biryawaza	biriya + waza	priya + vaja	«Vincitore-caro / che ha vittoria preziosa»
Sattiwaza	satti + waza	satya + vaja	«Che ha il premio della verita'»
Tushratta	tu + shratta	*tvesha + ratha	«Dal carro potente»
Artadama	arta + dama	rta + dama	«Casa dell'ordine cosmico»
Artamanya	arta + manya	rta + manya	«Che pensa all'ordine»
Shiwini	shi + wini	*surya-? / Mitra?	Divinita' solare mitannica
Shattara	shatta + ra	*ashta + ra	«Ottava» o «ricca»
Purra	pur + ra	*puru	«Abbondante / molto»

Appendice EE — Modelli matematici di filogenesi integrata linguistica-genomica: stato dell'arte 2025

L'integrazione formale di dati linguistici (cognate sets, morfologia comparata) e dati genomici (aDNA, f-statistics) in un unico modello probabilistico e' la frontiera metodologica piu' avanzata dell'Indo-europeistica quantitativa. Presentiamo qui un'introduzione ai principali approcci proposti nella letteratura recente.

L'approccio di Currat et al. (2023) propone di trattare la filogenia linguistica come un processo condizionato sulla storia demografica delle popolazioni, specificando un modello generativo congiunto:

$$P(L, G \mid T, \text{Theta}) = P(L \mid T_{lang}, \text{Theta}_{lang}) * P(G \mid T_{demo}, \text{Theta}_{demo}) * P(T_{lang} \mid T_{demo})$$

dove L e' la matrice dei dati linguistici (cognate sets), G e' la matrice dei dati genomici (f4-statistics e frequenze alleliche), T_lang e' l'albero linguistico, T_demo e' la storia demografica delle popolazioni, Theta_lang sono i parametri del modello di evoluzione linguistica, Theta_demo sono i parametri demografici, e P(T_lang | T_demo) e' il modello di correlazione tra struttura linguistica e struttura demografica.

Il termine critico e' P(T_lang | T_demo): come specificare la distribuzione a priori sulla concordanza tra albero linguistico e albero demografico? In assenza di sostituzione linguistica (ogni migrazione porta la propria lingua), i due alberi sarebbero identici. In presenza di sostituzione, possono divergere. Una specificazione ragionevole e' un modello di concordanza con parametro epsilon (0 = concordanza perfetta, 1 = indipendenza):

$$P(T_{lang} \mid T_{demo}, \text{epsilon}) \propto \exp(-\text{epsilon} * d_{topologica}(T_{lang}, T_{demo}))$$

dove d_topologica e' la distanza topologica tra i due alberi (ad esempio, il Robinson-Foulds distance). Con epsilon = 0 (concordanza perfetta), ogni divisione nell'albero demografico corrisponde a una divisione nell'albero linguistico; con epsilon grande, i due alberi sono quasi indipendenti. La stima bayesiana di epsilon dai dati IE produce valori nell'intervallo 0.1-0.3, indicando alta ma non perfetta concordanza tra storia demografica e storia linguistica — consistente con la presenza di alcune sostituzioni linguistiche (bilinguismo seguito da adozione della lingua IE) che non corrispondono a un cambiamento genomico discontinuo.

Un approccio alternativo, proposto da Heggarty e Renfrew (2014) e formalizzato da Bouckaert et al. (2022), utilizza un modello di diffusione spaziale bayesiana (Phylogeography) che inferisce simultaneamente l'albero linguistico e le traiettorie geografiche di ciascun ramo come processi di movimento browniano su una superficie di resistenza (che tiene conto di barriere geografiche, distanza e topografia). In questo framework, la datazione dei nodi dell'albero e' vincolata sia

dai dati linguistici (calibrazione da attestazioni storiche) sia dalla struttura spaziale (la velocita' di diffusione linguistica e' limitata da un parametro biologicamente ispirato). La combinazione di questi due tipi di vincoli produce stime di date piu' robuste di quelle ottenibili solo dai dati linguistici.

Appendice FF — Cronologia sinottica definitiva e indice bibliografico critico della Parte Quinta

La tavola cronologica sinottica seguente riassume, in forma comparativa, le principali tappe della storia indoeuropea dalla formazione del PIE al periodo storico documentato, integrando le evidenze di linguistica, archeologia e paleogenomica discusse nell'intera monografia.

Data (a.C.)	Evento linguistico/culturale	Cultura archeologica	Evidenza genomica chiave	Incertezza
7000-5500	Formazione delle pop. CHG e EHG pre-PIE	Cac.-racc. del Caucaso, Mesolitico steppa	Kotias Klde, Satsurblia (CHG)	Bassa
5500-4700	Proto-PIE: contatto EHG-CHG nelle steppe	Samara eneol., Khvalynsk incipiente	Yuzhniy Oleniy Ostrov, Khvalynsk precoce	Media
4700-4200	PIE arcaico: Khvalynsk, separaz. proto-anatolica?	Khvalynsk, Sredny Stog incipiente	Khvalynsk ~50% EHG+~43% CHG	Media
4500-4000	Separazione del proto-anatolico verso Balcani	Suvorovo-Novodanilovka (Delta Danubio)	Segnale steppico in Tracia/Romania BA	Alta
4000-3500	PIE nucleare (non-anatolico): Sredny Stog/Yamnaya incip.	Tardo Sredny Stog, Repin	Profilo Yamnaya ancestrale	Media
3500-2800	Yamnaya: espansione demografica nelle steppe	Yamnaya (Volga-Danubio)	Yamnaya: 52% EHG + 43% CHG	Bassa
3300-2800	Proto-anatolico: arrivo in Anatolia (via Balcani?)	Transizione EBA in Anatolia	Bassa comp. steppa in Anatolia BA	Alta

Data (a.C.)	Evento linguistico/culturale	Cultura archeologica	Evidenza genomica chiave	Incertezza
2900-2500	Proto-IE europeo: espansione Corded Ware	Corded Ware (Europa centrale)	CW: ~40% EHG+CHG, R1a dominante	Bassa
2500-2000	Proto-Celtico-Germanico: Bell Beaker	Bell Beaker (Pan-europeo)	BB: ~50% steppa UK, R1b-P312	Bassa
2200-1800	Proto-Indo-iranico: Sintashta	Sintashta (Urali meridionali)	Sintashta: ~75% steppa, R1a-Z93	Bassa
2000-1500	Proto-Indo-ariano: dispersione Andronovo-BMAC	Andronovo, BMAC transizione	Swat Valley: comp. steppa ~35%	Media
1600-1380	Mitanni-ariano: IE nel Vicino Oriente (teonimi)	Mitanni (Siria settentrionale)	Assente (solo linguistica)	Media (solo linguistica)
1650-1180	Ittita: IE anatolico documentato	Impero ittita (Hattusa)	Non disponibile (no aDNA sufficiente)	Bassa (lin.) - Alta (genomica)
1500-1000	Vedico: primo ramo IE con corpus letterario	Periodo Rigvedico	Inferita da aDNA Swat + gradiente moderno	Media
1200-800	Collasso Bronzo Tardo: redistribuzione IE	Popoli del Mare, transizione BA-FA	aDNA limitato dal periodo	Alta
800-200	IE documentato: Greco, Latino, Celtico, Iranico	Eta' del Ferro	Popolazioni moderne come proxy	Bassa (storica)

La monografia «Le Origini dei Paleoindoeuropei» conclude qui il suo percorso in cinque parti. L'arco che va dalla formazione delle popolazioni CHG ed EHG nelle steppe dell'Eurasia settentrionale — seimila anni fa, in un'epoca senza scrittura e senza Stato — fino alle lingue di un miliardo di parlanti contemporanei, e' uno degli archi narrativi piu' ampi e meglio documentati della storia umana. La scienza non ha ancora detto l'ultima parola: le scoperte dei prossimi dieci anni di paleogenomica, di analisi bayesiana e di nuovi scavi archeologici porteranno certamente precisioni, revisioni e sorprese. Ma il quadro fondamentale — la patria steppica, l'espansione demografica, la dispersione linguistica — e' ora abbastanza solido da resistere anche agli sviluppi piu' imprevisti.

INTRODUZIONE ALLE APPENDICI GENERALI

Le presenti Appendici Generali raccolgono i materiali di riferimento che accompagnano l'intera monografia «Le Origini dei Paleoindoeuropei» nelle sue cinque parti, complementando gli apparati critici già presenti al termine di ciascuna parte (Appendici A-Q per la Parte Prima-Terza, R-Z per la Parte Quarta, AA-FF per la Parte Quinta). Mentre quelle appendici sono tematicamente organizzate per parte, le presenti coprono trasversalmente l'intero arco della monografia.

L'organizzazione segue criteri funzionali: l'Appendice I contiene l'indice analitico completo dei termini tecnici; l'Appendice II raccoglie la tavola comparativa dei sistemi di ricostruzione del PIE nelle diverse scuole; l'Appendice III fornisce il corpus sistematico delle radici PIE con attestazioni in tutte le lingue figlie; l'Appendice IV elenca tutti i siti aDNA citati nella monografia con i relativi riferimenti; l'Appendice V offre la cronologia sinottica generale che integra tutti e cinque i capitoli della monografia; l'Appendice VI raccoglie la tavola completa delle culture archeologiche citate; l'Appendice VII fornisce il glossario integrato dei termini tecnici di linguistica, genomica e metodi matematici.

APPENDICE I — INDICE ANALITICO DEI TERMINI TECNICI

Il seguente indice raccoglie i termini tecnici di maggiore frequenza o importanza nell'intera monografia, con indicazione della parte e del capitolo in cui vengono introdotti e/o discussi in modo sistematico. Per i termini di uso comune nella linguistica storica si rimanda al Glossario (Appendice VII). I termini in maiuscoletto indicano forme proto-indoeuropee ricostruite; i termini in corsivo indicano nomi tecnici di metodi, culture o concetti.

Termine	Definizione sintetica	Parte/Capitolo principale
ABBA-BABA (D-statistic)	Test di flusso genico tra popolazioni basato su quattro campioni e pattern di alleli ancestrali/derivati	Parte I / Cap. II.3.2
Ablaut (apofonia)	Alternanza vocalica morfologicamente condizionata nel PIE (es. *e/*o/*zero)	Parte II / Cap. III.2.1
aDNA (ancient DNA)	DNA estratto da resti scheletrici di individui preistorici o storici	Parte I / Cap. II.0
ADMIXTURE	Metodo di stima delle proporzioni di ascendenza ancestrale attraverso k componenti latenti	Parte I / Cap. II.2
Afanasievo	Cultura steppica dell'Asia centrale (3300-2500 a.C.) interpretata come precursore del tocarico	Parte III / Cap. IX.2
Andronovo	Complesso culturale delle steppe asiatiche (2100-1400 a.C.), associato alla dispersione indo-iraniana	Parte II / Cap. IV.1 e VIII.4
ARG (Ancestral Recombination Graph)	Generalizzazione del coalescente che include eventi di ricombinazione	Parte V / App. EE
Arcaismo	Forma o tratto linguistico che preserva uno stadio antico della lingua proto-madre	Parte II / Cap. III.1.2
BEAST	Software bayesiano per filogenesi molecolare e linguistica (Bayesian Evolutionary Analysis Sampling Trees)	Suppl. Mat. / Cap. M2
Bell Beaker	Complesso culturale pan-europeo (2800-1800 a.C.) associato all'espansione del ramo centum occidentale	Parte II / Cap. IV.2
Bipartizione (split)	Divisione dell'insieme delle lingue in due sottoinsiemi, definita da un ramo interno dell'albero	Suppl. Mat. / Cap. M5

Termine	Definizione sintetica	Parte/Capitolo principale
Calibrazione (bayesiana)	Vincolo temporale su un nodo dell'albero filogenetico basato su attestazioni storiche	Suppl. Mat. / Cap. M2.3
CHG (Caucasus Hunter-Gatherer)	Componente ancestrale dei cacciatori-raccoglitori del Caucaso, co-fondatrice dello Yamnaya	Parte I / Cap. I.3.3
Coalescenza	Processo di risalita nel tempo dei lineage genomici verso un antenato comune	Parte I / Cap. II.0
Cognate	Forma lessicale che discende da un comune protolessema della lingua proto-madre	Parte II / Cap. III.1
Corded Ware (CW)	Cultura dell'Europa centrale (2900-2350 a.C.) che rappresenta la prima ondata di espansione IE documentata genomicamente	Parte I / Cap. II.1.2
Dendrogramma	Rappresentazione grafica ad albero delle relazioni filogenetiche o delle distanze lessicali	Suppl. Mat. / Cap. M2
Diffusione demica	Espansione di una cultura o di una lingua attraverso il movimento fisico di popolazioni	Parte I / Cap. I.2
Dollo (modello di)	Modello di sostituzione cognatica che permette perdite multiple ma acquisizioni uniche	Suppl. Mat. / Cap. M2.2
EEF (Early European Farmer)	Componente ancestrale degli agricoltori anatolici giunti in Europa nel Neolitico	Parte I / Cap. I.2.1
EHG (Eastern Hunter-Gatherer)	Componente ancestrale dei cacciatori-raccoglitori orientali, co-fondatrice dello Yamnaya	Parte I / Cap. I.2.1
Filogenia bayesiana	Metodo di stima dell'albero evolutivo attraverso l'inferenza bayesiana sulla verosimiglianza dei dati	Suppl. Mat. / Cap. M2
Fisher-KPP	Equazione di reazione-diffusione che modella la velocita' di avanzamento di una popolazione	Parte III / Cap. XIV.2.1
F-statistiche	Famiglia di statistiche (F2, F3, F4) per misurare la condivisione di deriva genetica tra popolazioni	Parte I / Cap. II.3
Glottocronologia	Metodo di datazione della separazione linguistica basato sul tasso di perdita del lessico di base	Suppl. Mat. / Cap. M1
Isoglossa	Confine geografico che separa aree con caratteristiche linguistiche diverse	Parte III / Cap. VII.3
Kurgan	Tumulo funerario delle steppe eurasiatiche, caratteristico delle culture PIE	Parte I / Cap. I.3

Termine	Definizione sintetica	Parte/Capitolo principale
Laringale	Fonema PIE ricostruito (H1, H2, H3) la cui esistenza fu prevista da de Saussure nel 1879 e confermata dall'ittita	Parte II / Cap. III.1.2
Legge di Grimm	Spostamento consonantico sistematico che distingue il ramo germanico dalle altre lingue IE	Parte II / Cap. III.1.3
Lee-Hasegawa (modello)	Modello di diffusione-reazione per la sostituzione linguistica, analogo alla Fisher-KPP	Parte III / Cap. XIV.4
Levenshtein (distanza di)	Numero minimo di operazioni elementari per trasformare una stringa in un'altra	Suppl. Mat. / Cap. M1.3
Lignaggio	Sequenza genomica tracciata attraverso il tempo verso un antenato comune	Parte I / Cap. II.6
Lineare B	Sillabario miceneo usato per scrivere il greco miceneo (1500-1200 a.C.)	Parte IV / App. V
Lituano	Lingua baltica considerata la piu' arcaica tra le lingue IE viventi	Parte IV / Cap. XVI.2.1
Maykop	Cultura del Caucaso settentrionale (3700-3000 a.C.) che introdusse il contributo CHG nello Yamnaya	Parte I / Cap. I.3.3
Mitanni-ariano	Varietà di proto-indo-ariano documentata nel regno di Mitanni (XVI-XIV sec. a.C.)	Parte V / Cap. XXVII
Nodo (albero)	Punto di biforcazione in un albero filogenetico, corrispondente a un evento di separazione	Suppl. Mat. / Cap. M2
PCA (Principal Component Analysis)	Metodo di riduzione della dimensionalità applicato ai dati genomici per visualizzare la struttura di popolazione	Parte I / Cap. II.1.3
PIE	Proto-Indo-Europeo, lingua proto-madre ricostruita della famiglia indoeuropea	Tutte le parti
qpAdm	Metodo per stimare le proporzioni di miscela tra popolazioni usando le statistiche f	Parte I / Cap. II.4
Rake (topologia a)	Configurazione dell'albero con multiple radiazioni quasi-simultanee, difficile da risolvere filogeneticamente	Parte III / Cap. VII.1.2
ROH (Runs of Homozygosity)	Tratti genomici omozigoti che indicano inbreeding o storia di piccola popolazione	Parte I / Cap. II.5
Satem/Centum	Divisione tradizionale delle lingue IE in due gruppi basata sul trattamento delle velari PIE	Parte III / Cap. VII.2

Termine	Definizione sintetica	Parte/Capitolo principale
Sintashta	Cultura delle steppe meridionali (2200-1800 a.C.), identificata come culla del proto-indo-iranico	Parte II / Cap. IV.1.1
Sprachbund	Area linguistica in cui lingue non imparentate condividono tratti strutturali per contatto	Parte III / Cap. VII.3
Urheimat	Patria originale dei Proto-Indoeuropei, oggetto del dibattito centrale della monografia	Parte I / Cap. I.1
Vedico	Stadio piu' arcaico del proto-indo-ariano, preservato nel Rigveda (c. 1500-1200 a.C.)	Parte III / Cap. VIII.1
WHG (Western Hunter-Gatherer)	Componente ancestrale dei cacciatori-raccoglitori europei occidentali, presente in Europa prima del Neolitico	Parte I / Cap. I.2.1
Yamnaya	Cultura delle steppe pontico-caspiche (3500-2800 a.C.), ritenuta principale vettore dell'espansione IE	Parte I / Cap. I-II

Tabella I.1 — Indice analitico dei principali termini tecnici della monografia

APPENDICE II — TAVOLA COMPARATIVA DELLE SCUOLE DI RICOSTRUZIONE DEL PIE

La ricostruzione del proto-indoeuropeo non è un'impresa monolitica: diverse tradizioni scolastiche hanno prodotto sistemi di ricostruzione parzialmente incompatibili, con differenze significative nella fonologia, nella morfologia e perfino nel vocabolario ricostruito. La tavola seguente confronta le principali scuole nella loro posizione sui problemi teorici fondamentali.

Problema teorico	Scuola tradizionale (Brugmann-Meillet)	Scuola laringale (Kurilowicz-Cowgill)	Scuola americana (Watkins-Fortson)	Glottalica (Gamkrelidze-Ivanov)	Tradizione tedesca moderna (Clackson-Fortson)
Numero di laringali PIE	*H1, *H2, *H3 (tre)	*H1, *H2, *H3 (tre), piu' varianti	*H1, *H2, *H3 (tre, consenso)	Rianalisi parziale (glottalizzazione)	Tre, con distinzioni funzionali
Sistema delle occlusive	Due serie sorde (*p,*t,*k) + sonore (*b,*d,*g) + sonoraspirate	Idem, con rianalisi delle laringali	Consenso tradizionale, tre serie	Teoria glottalica: ejettive invece di sonore semplici	Tre serie, con dibattito su *b-
Problema di *b-	Fonema raro ma esistente	Raro, spiegato come laringale	Raro, probabilmente deriva da *bh-	Non-problematico nella teoria glottalica	Raro, segnale di prestito
Vocalismo	Cinque vocali (*a, *e, *i, *o, *u)	Solo *e e *o di base; *a da laringale	Solo *e, *o originarie; *a marginale	Come laringalisti	Due vocali di base (*e, *o)
Ordine delle parole PIE	SOV con flessibilita'	SOV tendenziale	SOV con enclitiche in seconda posizione	SOV rigido	SOV, con legge di Wackernagel
Data di separazione stimata	Non formalizzata (pre-bayesiana)	Non formalizzata	5000-6000 a.C. (aggiornamenti recenti)	Analoga alla tradizione	4500-6000 a.C.
Posizione dell'anatolico	Ramo basale arcaico	Ramo basale con laringali conservate	Ramo basale, PIE-anatolico (PAIe)	Ramo basale	Ramo basale, «indo-ittita» dibattuto

Tabella II.1 — Posizioni delle principali scuole sui problemi teorici fondamentali del PIE

Una seconda tavola compare le sistematizzazioni dei riflessi delle laringali nelle principali lingue figlie IE, secondo la ricostruzione standard a tre laringali (*H1, *H2, *H3). Questa tavola integra la discussione nel Cap. III.1.2 della Parte Seconda.

Contesto	Proto-IE	Ittita	Greco	Latino	Sanscrito	Avestico	Lituano	Armeno
*H1 tra consonanti	*H1e-	he-	e-	e-	a-	a-	e-	e-
*H2 tra consonanti	*H2e-	ha-	a-	a-	a-	a-	a-	a-
*H3 tra consonanti	*H3e-	ha-	o-	a-	a-	a-	a-	o-
*VH1C (vocale prima)	Allungamento	h-conservato	Allungamento	Allungamento	Allungamento	Allungamento	Allungamento	Allungamento
*VH2C (vocale prima)	Allungamento	h-conservato	Allungamento	Allungamento	Allungamento	Allungamento	Allungamento	Allungamento
*H2 + *e + R	*H2er-	har-	ar-	ar-	ar-	ar-	ar-	ar-
*H3 + *e + R	*H3er-	har-	or-	ar-	ar-	ar-	ar-	or-

Tabella II.2 — Riflessi delle laringali PIE nelle lingue figlie principali

APPENDICE III — CORPUS SELEZIONATO DELLE RADICI PIE CON ATTESTAZIONI

Il seguente corpus raccoglie 80 radici PIE di particolare importanza per la monografia — sia per la loro centralità nel lessico culturale ricostruito (cfr. Cap. III.3 della Parte Seconda) sia per le discussioni di teoria fonologica e morfologica nelle Parti Seconda-Quinta. Per ciascuna radice sono indicate: la forma ricostruita standard (con notazione laringale), il significato ricostruito, e i principali riflessi nelle lingue figlie. La notazione segue la tradizione di Pokorny (IEW) aggiornata secondo LIV2 (Rix et al. 2001).

Le radici sono organizzate per campo semantico: (A) lessico del corpo e della parentela, (B) lessico degli animali e dell'ambiente, (C) lessico del carro e della tecnologia, (D) lessico del potere e della societa', (E) lessico del sacro e della cosmologia, (F) lessico verbale fondamentale.

A. Lessico del corpo e della parentela

Forma PIE	Significato	Greco	Latino	Sanscrito	Gotico	Lituano	Notazione LIV2
*H2ner-	Uomo, virile	aner	Ner (nome)	nar-	—	noras	IEW 765
*pHter-	Padre	pater	pater	pitar-	fadar	tevas	IEW 829
*maHter-	Madre	meter	mater	matar-	aiths	motina	IEW 700
*bHraHter-	Fratello	phrater	frater	bhratar-	brothar	brolis	IEW 163
*swesor-	Sorella	eor (eleo)	soror	svasar-	swistar	sesuo	IEW 1051
*suHnus	Figlio	—	—	sunu-	sunus	sunus	IEW 914
*H1okw-	Occhio	osse (dual.)	oculus	akshi	augo	akis	IEW 775
*H2ows-	Orecchio	ous	auris	us-	auso	ausis	IEW 785

Forma PIE	Significato	Greco	Latino	Sanscrito	Gotico	Lituano	Notazione LIV2
*nHsus	Naso	naso (metat.)	nasus	nasa-	naso	nosis	IEW 755
*H1dont-	Dente	odous	dens	dant-	tunthus	dantis	IEW 287
*gwhen-	Donna, moglie	gyne	—	jani-	qino	—	IEW 473
*sneusH-	Nuora	nyes	nurus	snusha-	—	snasa	IEW 977

Tabella III.A — Lessico del corpo e della parentela: radici PIE selezionate

B. Lessico degli animali e dell'ambiente

Forma PIE	Significato	Greco	Latino	Sanscrito	Gotico	Lituano
*H1ekwos	Cavallo	hippos	equus	asva-	aihvs	asva
*gwous	Bovino (bue/mucca)	bous	bos	gaus	auhsa	guovs
*H2owi-s	Pecora	ois	ovis	avi-	awi	avis
*suH-s	Maiale (cinghiale)	hys	sus	sukara-	—	sunis
*kwon-	Cane	kyon	canis	svan-	hunds	suo
*kapro-	Capra	kapros	caper	—	haira	kairys
*wornis	Aquila, uccello rapace	ornis	—	vaynas	—	erelis
*pisk-	Pesce	—	piscis	—	fisks	zuvis
*H2erH3-	Arare	aroo	arare	—	arjan	arti
*H2egros	Campo	agros	ager	ajra-	akrs	agras
*deru-	Albero, legno, quercia	drys	larix (parz.)	daru-	triu	derva
*H2ep-	Acqua (fiume)	apee	amnis (parz.)	ap-	—	upe

Forma PIE	Significato	Greco	Latino	Sanscrito	Gotico	Lituano
*sneygwh-	Neve	nipha	nix	sneha-	snaiws	sniegas

Tabella III.B — Lessico degli animali e dell'ambiente: radici PIE selezionate

C. Lessico del carro, della tecnologia e della guerra

Il lessico del carro e della ruota e' di importanza fondamentale per la datazione dell'espansione PIE: la presenza di questo lessico in tutte le principali sottofamiglie IE (con la parziale eccezione del ramo anatolico) indica che la ruota era conosciuta dai parlanti del PIE nucleare, e poiche' la ruota e' attestata archeologicamente a partire dal 3500-3400 a.C. nelle steppe del Ponto-Caspio, questo costituisce un argomento forte per una data di dispersione del PIE non anteriore al 3500 a.C. (cfr. Cap. III.3.1 della Parte Seconda).

Forma PIE	Significato	Greco	Latino	Sanscrito	Antico alto-tedesco	Lituano
*kwekwlo-	Ruota, cerchio	kyklos	—	cakra-	hweohl	—
*H2eks-	Asse del carro	axon	axis	aksha-	ahsa	asis
*weg-h-	Trasportare sul carro	—	vehere	vah-	wegan	vegti
*H2iHHes-	Rame, metallo	aes (lat.)	aes	ayas-	er	—
*kowHro-	Gregge, bestiame come ricchezza	—	—	kula-	—	—
*H2ewdHH-	Arma, strumento	autos (parz.)	ausum	udhas-	—	—
*perkwu-	Quercia sacra / Forza del fulmine	perke (dor.)	Quercus	Parjanya	—	Perkunas
*gwerH-	Forgiare, fabbricare	—	—	grnati	—	—

Forma PIE	Significato	Greco	Latino	Sanscrito	Antico alto-tedesco	Lituano
*kwey-	Pagare, punire	tino	poena	ci-	—	kaina
*spek-	Spiare, osservare	skeptomai	specio	spasati	—	spegti

Tabella III.C — Lessico del carro, della tecnologia e della guerra

D. Lessico del potere, della societa' e della cosmologia

Forma PIE	Significato	Greco	Latino	Sanscrito	Avestico	Lituano
*H3regHs	Re, sovrano	— (iranico)	rex	raj-	razah-	—
*H2eyHs-	Vita, forza vitale	aion	aetas	ayu-	ayuh-	amzius
*H1es-	Essere (copula)	esti	esse	asti	asti	esti
*deywo-	Divo, celeste	Zeus (gen.)	deus	deva-	daeva-	dievas
*H2su-	Buono, giusto	—	—	su-	hu-	—
*wlkwos	Lupo (figura mitica)	lykos	lupus	vrka-	vhrka-	vilkas
*H2ner-	Eroe, guerriero	aner	—	nar-	nar-	vyras
*koryos	Banda guerriera giovanile	koira	—	senas	haena	—
*sweH2dH-	Giurare, promettere	—	suadeo	svadata-	—	—
*H1lewdH-	Popolo libero	laos	—	—	—	liaudis
*dyewpH2ter-	Padre celeste (Zeus-Pater)	Zeus pater (Giove)	Iuppiter	Dyauspitar	—	—

Forma PIE	Significato	Greco	Latino	Sanscrito	Avestico	Lituano
*H2uHros	Signore, protettore	ouros	—	vara-	ahura-	—

Tabella III.D — Lessico del potere, della societa' e della cosmologia PIE

APPENDICE IV — SITI aDNA CITATI NELLA MONOGRAFIA

La seguente tabella raccoglie tutti i principali siti archeologici da cui provengono campioni aDNA discussi nella monografia, con indicazione della cultura, dell'aplogruppo Y prevalente, del componente ancestrale dominante, e dello studio di riferimento. La tabella e' organizzata cronologicamente all'interno di macroaree geografiche.

A. Steppe pontico-caspiche e regioni adiacenti

Sito / Cultura	Data (a.C. approx.)	Localita'	Aplogruppo Y	Componente ancestrale	Studio chiave
Khvalynsk eneolitico	4700-4200	Volga medio, Russia	R1b, Q	~50% EHG + ~43% CHG	Mathieson et al. 2015
Sredny Stog	4500-3500	Ucraina orientale	R1a, R1b	Proto-Yamnaya (transizione)	Anthony et al. 2022
Maykop	3700-3000	Caucaso nord-occidentale	Vari	CHG-dominante + Iran	Wang et al. 2019
Yamnaya (Volga)	3500-2800	Volga-Don, Russia	R1b-M269	52% EHG + 43% CHG	Haak et al. 2015
Yamnaya (Ucraina)	3500-2800	Ucraina, Moldavia	R1b-L23	Simile al Volga	Haak et al. 2015
Catacomb (Ucraina)	2800-2200	Ucraina meridionale	R1b, R1a	Simile Yamnaya + EEF lieve	Mathieson et al. 2015
Sintashta	2200-1800	Urali meridionali, Russia	R1a-Z93	~75% steppa	Allentoft et al. 2015
Andronovo	2100-1400	Kazakistan-Siberia	R1a-Z93	Simile Sintashta	Allentoft et al. 2015
Srubna / Srubnaya	2100-1500	Steppa pontica	R1a, R1b	Steppa con lieve EEF	Mathieson et al. 2015
Afanasievo	3300-2500	Altai, Siberia merid.	R1b	Simile Yamnaya	Damgaard et al. 2018

Tabella IV.A — Principali siti aDNA delle steppe eurasiatiche citati nella monografia

B. Europa durante l'espansione IE (Corded Ware e Bell Beaker)

Sito / Cultura	Data (a.C. approx.)	Area geografica	Aplogruppo Y	Comp. steppa (%)	Studio chiave
Corded Ware (Germania)	2900-2350	Europa centrale	R1a-M417	~45-60%	Haak et al. 2015
Corded Ware (Polonia)	2900-2350	Polonia	R1a	~40-55%	Mathieson et al. 2015
Bell Beaker (Iberia)	2500-1800	Iberia	R1b-P312	~40-60%	Olalde et al. 2018
Bell Beaker (Gran Bretagna)	2450-2000	Isole Britanniche	R1b-L21	~50-90%	Olalde et al. 2018
Bell Beaker (Paesi Bassi)	2500-1900	Europa nord-occidentale	R1b	~50-70%	Olalde et al. 2018
Cultura di Unetice	2300-1600	Boemia, Germania	R1b, R1a	~40-55%	Allentoft et al. 2015
Cultura di Unetice (Sassonia)	2300-1600	Germania centrale	R1b	~50%	Allentoft et al. 2015
Cultura dei Campi d'Urne	1300-800	Europa centro-oc.	R1b	Simile BB + variazione	Knipper et al. 2017
Hallstatt	800-500	Austria, Svizzera	R1b	Base BB + variazione	Cunliffe 2019
La Tene / Celtico	500-50	Europa centro-oc.	R1b-L21	Base BB	Vari studi

Tabella IV.B — Siti aDNA dell'Europa nell'eta' del Bronzo durante l'espansione IE

APPENDICE V — CRONOLOGIA SINOTTICA GENERALE DELL'INTERA MONOGRAFIA

La seguente cronologia integra le tavole parziali presenti nelle appendici delle cinque parti (Appendice Q-Parte Terza, Appendice Z-Parte Quarta, Appendice FF-Parte Quinta), offrendo una visione d'insieme che copre l'intero arco dalla formazione delle popolazioni ancestrali alla fine dell'eta' del Bronzo. I periodi di alta incertezza cronologica sono indicati con asterisco (*).

Periodo (a.C.)	Sviluppi demografici/genetici	Cultura archeologica	Stadio linguistico	Incertezza
8000-7000	Presenza WHG e EHG in Europa e steppe; CHG nel Caucaso	Mesolitico europeo	Pre-PIE: lingue dei cacciatori-raccoglitori	Alta
7000-5500	Arrivo EEF dalla Anatolia in Europa; mescolanza con WHG	Neolitico LBK, Impressa, Cardiale	Pre-PIE / lingue neolitiche anatoliche	Media
5500-4700	Contatto EHG-CHG nelle steppe del Volga	Samara eneolitico, proto-Khvalynsk	Protopie: fase pre-laringale?*	Alta
4700-4200	Cultura Khvalynsk: sintesi EHG (50%) + CHG (43%)	Khvalynsk, proto-Sredny Stog	PIE arcaico: sistema laringale formato	Media
4500-3900	Cultura Sredny Stog: precursore Yamnaya	Sredny Stog, Skelya	PIE nucleare (pre-anatolico?)*	Media-Alta
4000-3700	Cultura Maykop nel Caucaso nord-occ.	Maykop	Contatti PIE-Caucasico?*	Alta
3700-3500	Transizione verso Yamnaya nel Volga-Don-Dnjepr	Proto-Yamnaya (Repin)	PIE nucleare: carro gia' nel lessico	Media
3500-2800	Yamnaya: espansione nelle steppe (Volga-Danubio)	Yamnaya classico	PIE nucleare: dispersione inizia	Bassa
3300-2500	Espansione prototocarica via Afanasievo in Asia centrale	Afanasievo (Altai)	Proto-Tocarico: ramo precoce	Media
3300-2800	Migrazione proto-anatolica verso Balcani e Anatolia?	Suvorovo-Novodanilovka	Proto-Anatolico: separazione*	Alta

Periodo (a.C.)	Sviluppi demografici/genetici	Cultura archeologica	Stadio linguistico	Incertezza
2900-2500	Migrazione Yamnaya in Europa centrale: Corded Ware	Corded Ware (Europa centrale)	Proto-IE non-anatolico: radice europea	Bassa
2800-2000	Espansione Bell Beaker da Iberia verso Europa	Bell Beaker (pan-europeo)	Proto-Celtico-Italico-Germanico incip.*	Bassa
2500-2000	IE-ficazione delle Isole Britanniche (sostituzione quasi totale)	Bell Beaker Gran Bretagna	Proto-Celtico insulare	Bassa
2200-1800	Sintashta: proto-indo-iranico con carri da guerra	Sintashta, Petrovka	Proto-Indo-Iranico definito	Bassa
2100-1500	Espansione Andronovo in Asia centrale	Andronovo	Proto-Indo-Ariano / Proto-Iranico in sep.	Bassa-Media
2000-1700	Espansione greco in Grecia (Egeo)	EBA III, MH	Proto-Greco in Egeo	Media
1800-1500	Espansione Indo-Ariana verso Sud-Asia via BMAC	BMAC transizione, Swat Valley	Proto-Indo-Ariano nel Subcontinente	Media
1650-1380	Mitanni-ariano nel Vicino Oriente (Siria-Mesopotamia)	Mitanni	Mitanni-ariano: dialetto proto-IA	Media
1650-1200	Ittita documentato: Impero ittita	Impero ittita (Hattusa)	Anatolico documentato	Bassa
1500-1100	Vedico: primo testo letterario IE (Rigveda)	Panjab / Cultura vedica	Vedico arcaico	Media
1500-1100	Miceneo documentato (Lineare B)	Micene, Cnosso, Pilo	Greco miceneo documentato	Bassa
1200-800	Collasso del Bronzo Tardo: ridistribuzione IE	Popoli del Mare, FA	IE in espansione secondaria	Alta
800-300	Eta' del Ferro: dispersione finale di tutti i rami IE	Periodo storico	IE pienamente documentato in tutte le rami	Bassa

Tabella V.1 — Cronologia sinottica generale: dall'8000 al 300 a.C.

APPENDICE VI — TAVOLA DELLE CULTURE ARCHEOLOGICHE CITATE

La seguente tavola riassume le principali culture archeologiche citate nell'intera monografia, con indicazione della datazione calibrata (a.C.), dell'area di distribuzione, delle caratteristiche diagnostiche principali, e della relazione con l'espansione IE secondo la letteratura corrente.

Cultura archeologica	Data (a.C.)	Area	Caratteristiche diagnostiche	Relazione con IE
Khvalynsk	4700-3800	Volga medio	Sepolture ocra, kurgan proto-forme, ceramica a fondo acuto	Precursore diretto dello Yamnaya; PIE arcaico
Sredny Stog	4500-3500	Ucraina or.	Sepolture in posizione supina; addomesticamento del cavallo	Transizione EHG-CHG verso Yamnaya
Maykop	3700-3000	Caucaso no.	Metallurgia avanzata; ceramica dipinta; sepolture kurgan ricche	Fonte del contributo CHG nello Yamnaya
Yamnaya	3500-2800	Steppe V-D	Kurgan a pozzetto; ocra rossa; mobilita' pastorale	Principale vettore dell'espansione PIE in Europa
Afanasievo	3300-2500	Altai-Yenis.	Simile a Yamnaya; ceramica a fondo ovoidale	Ramo proto-tocarico; migrazione precoce verso est
Corded Ware	2900-2350	Eur. centr.	Ceramica cordata; asce da battaglia; sepolture individuali	Prima ondata IE in Europa; proto-balto-slavo-germanico
Bell Beaker	2800-1800	Pan-europeo	Ceramiche campaniformi; punte di freccia; rame	Espansione IE verso ovest; proto-celtico-italico
Sintashta	2200-1800	Urali mer.	Carri da guerra a due ruote; cavalli sacrificali; fortezze	Culla del proto-indo-iranico
Andronovo	2100-1400	Kazak.-Sib.	Steppe asiatiche centrali; espansione a est e sud	Dispersione indo-ariana e iranica

Cultura archeologica	Data (a.C.)	Area	Caratteristiche diagnostiche	Relazione con IE
Srubna	2100-1500	Steppa pont.	Ceramica a larghe spalle; pastorizia mobile	Proto-iranico occidentale (sciti, sarmati anc.)
Unetice	2300-1600	Boh.-Ger.	Bronzo avanzato; sepolture ricche; commercio a lunga distanza	Transizione CW-BB; proto-germanico/italico
Terramare	1700-1150	Pianura Pad.	Villaggi su palafitte; ceramica terramare; bronzo	Proto-italico in Italia settentrionale
Campi d'Urne	1300-800	Eur. cent.	Cremazione sistematica; deposizioni in urne fittili	Proto-celtico continentale; transizione FA
Hallstatt	800-500	Eur. alp.	Sale; carri funerari; fibule; principi celto-hallstatt	Celtico documentato primo; La Tene successivo
BMAC (Bactria-Margiana)	2300-1700	Asia centr.	Citta' murate; ceramica sofisticata; lapis lazuli	Punto di transito indo-ariano verso il Subcontinente
Swat Valley	1800-1200	Pakistan	Necropoli con ricca ceramica; continuita' con Gandhara Grave	Primo aDNA con componente steppa nel Subcontinente
Mitanni	1600-1350	Siria sett.	Teonimi ariani; tecnica del cavallo (kikkuli-text)	Unico IE attestato nel Vicino Oriente antico (non-anatolico)

Tabella VI.1 — Culture archeologiche principali citate nella monografia

APPENDICE VII — GLOSSARIO INTEGRATO DEI TERMINI TECNICI

Il glossario raccoglie le definizioni dei termini tecnici di linguistica storica, paleogenomica e metodi matematici usati nella monografia. I termini segnalati con [L] appartengono al lessico della linguistica storica e comparativa; quelli con [G] alla paleogenomica e alle scienze della vita; quelli con [M] ai metodi matematici e statistici. Per i termini presenti nell'Indice Analitico (Appendice I) viene fornita una definizione piu' estesa.

A — Terminologia linguistica

Termine	Tipo	Definizione
Ablaut	[L]	Alternanza vocalica morfologicamente condizionata (*e/*o/*zero) nel PIE; anche «apofonia». Meccanismo centrale della morfologia nominale e verbale PIE.
Acculturazione linguistica	[L]	Adozione di elementi di una lingua straniera senza spostamento demografico; modalita' alternativa alla diffusione demica per la diffusione delle innovazioni linguistiche.
Alloglossia	[L]	Situazione in cui una lingua o varieta' e' geneticamente isolata (senza parenti noti) all'interno di un'area geografica; tipicamente riferita al basco nel contesto IE.
Anatolico	[L]	Sottofamiglia IE piu' arcaica, comprendente ittita, luvio, palaitico, licio, lidio; primo ramo a separarsi dal PIE.
Apofonia	[L]	Sinonimo di ablaut; alternanza sistematica di vocali in forme di una stessa radice.
Arcaismo	[L]	Forma o tratto linguistico che preserva uno stadio antico della proto-lingua senza innovazione.
Assimilazione	[L]	Processo per cui un fonema acquisisce caratteristiche di un fonema adiacente (regressiva: dalla destra; progressiva: dalla sinistra).
Bipartizione	[L]	Divisione dell'insieme delle lingue in due sottoinsiemi, definita da un ramo interno dell'albero filogenetico.
Calco	[L]	Prestito semantico: traduzione letterale di un composto o espressione straniera con materiale fonetico nativo.
Centum / Satem	[L]	Divisione tradizionale dei rami IE in due gruppi basata sul trattamento delle velari PIE: le lingue «centum» (latino, greco, celtico, germanico, tocario, anatolico) mantengono tre serie di velari; le lingue «satem» (indo-iranico, balto-slavo, armeno, albanese) fondono le velari palatali con le palatovelari.

Termine	Tipo	Definizione
Cognate	[L]	Forma lessicale in una lingua che discende da un comune protolessema con la forma corrispondente in un'altra lingua; opposto a «falso amico» (forme simili ma etimologicamente non imparentate).
Declinazione	[L]	Sistema di flessione nominale che indica la funzione sintattica di un sostantivo attraverso variazioni di desinenza (casi).
Diacronia	[L]	Studio di un sistema linguistico attraverso il tempo; opposto a «sincronia» (studio in un dato momento).
Glottocronologia	[L]	Metodo di datazione della separazione linguistica basato sul tasso di cambiamento del vocabolario di base; contestato ma ancora utilizzato nelle versioni bayesiane.
Isoglossa	[L]	Confine geografico che separa aree con caratteristiche linguistiche diverse (fonologiche, lessicali, morfologiche).
Laringale	[L]	Fonema PIE ricostruito (H1, H2, H3) postulato da Ferdinand de Saussure nel 1879 e confermato dall'ittita nel XX secolo; fondamentale per la ricostruzione del vocalismo PIE.
Legge fonetica	[L]	Corrispondenza regolare e sistematica tra fonemi di lingue imparentate, che riflette un cambiamento fonetico storico; es. Legge di Grimm, Legge di Verner.
Morfema	[L]	Unita' minima di significato grammaticale; le radici, i suffissi e le desinenze del PIE sono morfemi.
Prestito	[L]	Parola adottata da una lingua straniera, adattata fonologicamente o meno alla lingua ricevente.
Proto-lingua	[L]	Lingua ricostruita ipoteticamente come antenata di un gruppo di lingue imparentate; es. il PIE e' la proto-lingua della famiglia indoeuropea.
Radice	[L]	Morfema lessicale di base dal quale si formano le parole attraverso affissazione e derivazione.
Sprachbund	[L]	Area linguistica (Arealbund) in cui lingue non imparentate geneticamente condividono tratti strutturali per contatto prolungato.
Substrato	[L]	Influenza di una lingua precedente (perduta) su una lingua successiva nel medesimo territorio; es. substrato non-IE nelle lingue greche e latine.
Urheimat	[L]	Patria originale di un gruppo linguistico; per il PIE, il dibattito tra ipotesi steppica e ipotesi anatolica (e altre minori) e' il tema centrale della monografia.
Verner (legge di)	[L]	Corrispondenza sistematica (scoperta da Karl Verner nel 1875) che spiega le apparenti eccezioni alla Legge di Grimm attraverso il ruolo dell'accento PIE.

Tabella VII.A — Glossario dei termini di linguistica storica

B — Terminologia paleogenomica

Termine	Tipo	Definizione
aDNA	[G]	DNA antico (ancient DNA): DNA estratto da resti scheletrici preistorici o storici. Caratterizzato da frammentazione, deaminazione e rischio di contaminazione.
Aplogruppo	[G]	Insieme di linee genealogiche che condividono un antenato comune recente; nel contesto della monografia, gli aplogruppi del cromosoma Y (patrilineari) e del DNA mitocondriale (matrilineari) sono usati come marcatori demografici.
AMOVA	[G]	Analysis of Molecular Variance: metodo per partizionare la varianza genetica in componenti intra-gruppo e inter-gruppo.
Deaminazione	[G]	Processo di degradazione del DNA antico per cui la citosina si converte in uracile, producendo caratteristiche transizioni C->T nelle estremita' dei frammenti.
Deriva genetica	[G]	Cambiamento casuale delle frequenze alleliche in una popolazione di dimensione finita; piu' forte per popolazioni piccole.
F-statistiche	[G]	Famiglia di statistiche (F2, F3, F4) per misurare la condivisione di deriva genetica tra popolazioni; strumento centrale della paleogenomica moderna.
Flusso genico	[G]	Scambio di materiale genetico tra popolazioni attraverso la migrazione e la riproduzione.
Genotipo	[G]	Costituzione genetica di un individuo in uno o piu' loci genomici.
Haplogroup	[G]	Sinonimo di aplogruppo (dall'inglese).
IBD (Identity By Descent)	[G]	Segmenti genomici identici perche' ereditati da un antenato comune recente; lunghi segmenti IBD indicano parentela recente.
PCA	[G]	Principal Component Analysis: metodo di riduzione della dimensionalita' applicato ai dati genomici per visualizzare graficamente la struttura di popolazione.
qpAdm	[G]	Metodo computazionale per stimare le proporzioni di miscela tra popolazioni ancestrali usando le statistiche-f.
ROH	[G]	Runs of Homozygosity: lunghi tratti genomici omozigoti che indicano inbreeding o storia di piccola popolazione.
SNP	[G]	Single Nucleotide Polymorphism: variazione di un singolo nucleotide in una posizione genomica specifica.

Termine	Tipo	Definizione
Taglia efficace (Ne)	[G]	Dimensione di una popolazione ideale di Wright-Fisher che produce gli stessi effetti di deriva di una popolazione reale con struttura complessa.
WHG / EHG / CHG / EEF	[G]	Acronimi per le componenti ancestrali europee preistoriche: Western Hunter-Gatherer, Eastern Hunter-Gatherer, Caucasus Hunter-Gatherer, Early European Farmer.
Yamnaya	[G]	Cultura delle steppe pontico-caspiche (3500-2800 a.C.); genomicamente ~52% EHG + ~43% CHG; principale vettore della diffusione delle lingue IE in Europa.

Tabella VII.B — Glossario dei termini di paleogenomica

C — Terminologia matematica e statistica

Termine	Tipo	Definizione
ABC (Approximate Bayesian Computation)	[M]	Metodo di inferenza bayesiana che evita il calcolo esplicito della verosimiglianza attraverso la simulazione di statistiche sommario; usato per confrontare modelli demografici complessi.
Bayes factor	[M]	Rapporto delle verosimiglianze marginali di due modelli alternativi; misura quantitativa del supporto relativo dei dati per ciascun modello.
BEAST	[M]	Software bayesiano per la filogenesi molecolare e linguistica (Bayesian Evolutionary Analysis Sampling Trees); implementa modelli di orologio rilassato e calibrazioni nodali.
Catena di Markov	[M]	Processo stocastico in cui la distribuzione del passo successivo dipende solo dallo stato corrente, non dalla storia passata.
Coalescenza	[M]	Processo stocastico che descrive la risalita nel tempo dei lineage genomici verso un antenato comune; fondamentale per connettere i dati genomici con i parametri demografici.
Concordance factor (CF)	[M]	Proporzione di caratteri (o loci genomici) che supportano una data bipartizione dell'albero di consenso; misura di robustezza del segnale filogenetico.
Diffusione (coefficiente di)	[M]	Parametro che misura la mobilita' spaziale di una popolazione o di un'innovazione linguistica nella equazione di diffusione.
Distanza di Levenshtein	[M]	Numero minimo di operazioni elementari (inserzione, cancellazione, sostituzione) necessarie per trasformare una stringa in un'altra; usato per misurare la distanza fonetica tra cognati.

Termine	Tipo	Definizione
Fisher-KPP	[M]	Equazione di reazione-diffusione introdotta da Fisher (1937) e Kolmogorov-Petrovsky-Piskunov (1937); modella la velocita' di avanzamento di una popolazione in espansione.
Indice di Moran	[M]	Misura di autocorrelazione spaziale; valori vicini a 1 indicano raggruppamento spaziale, valori vicini a -1 indicano dispersione spaziale.
Verosimiglianza	[M]	P(dati | parametri): la probabilita' di osservare i dati dato il valore dei parametri del modello; quantita' centrale nell'inferenza statistica.
Variogramma	[M]	Funzione che misura come la varianza di un tratto aumenta con la distanza geografica tra i punti di osservazione; usato per quantificare la struttura spaziale dei dati linguistici.
MCMC	[M]	Markov Chain Monte Carlo: famiglia di algoritmi per campionare dalla distribuzione posteriore in inferenza bayesiana quando il calcolo analitico e' impraticabile.
Modello di Dollo	[M]	Modello di sostituzione cognatica che permette perdite multiple ma acquisizioni uniche di un tratto; corrisponde alla «legge di Dollo» dell'irreversibilita' evolutiva.
Orologio molecolare	[M]	Ipotesi secondo cui il materiale genetico (o cognatico) cambia a un tasso approssimativamente costante nel tempo; fondamento della datazione filogenetica.
Robinson-Foulds (distanza)	[M]	Misura della distanza topologica tra due alberi: conta il numero di bipartizioni presenti in uno ma non nell'altro; usata come metrica di concordanza nel modello integrato genomica-linguistica.

Tabella VII.C — Glossario dei termini matematici e statistici